

RIMS 共同研究「組合せ最適化セミナー」(第 22 回)

オンライン組合せ最適化とリグレット解析

東京大学大学院情報理工学系研究科

伊藤伸志

2025/07/22

講義の概要

- リグレット解析の枠組みとその応用
 - オンライン凸最適化問題とオンライン勾配降下法
 - Online-to-batch 変換
 - ミニマックス定理
 - エクスパート問題と Hedge アルゴリズム
- 正則化を用いたアルゴリズムとリグレット解析
 - オンライン鏡像降下法
 - 多腕バンディット問題と EXP3 アルゴリズム
- オンライン組合せ最適化
 - 完全情報フィードバック
 - セミバンディットフィードバック
 - バンディットフィードバック

講義の概要

- リグレット解析の枠組みとその応用
 - オンライン凸最適化問題とオンライン勾配降下法
 - Online-to-batch 変換
 - ミニマックス定理
 - エキスパート問題と Hedge アルゴリズム
- 正則化を用いたアルゴリズムとリグレット解析
 - オンライン鏡像降下法
 - 多腕バンディット問題とEXP3アルゴリズム
- オンライン組合せ最適化
 - 完全情報フィードバック
 - セミバンディットフィードバック
 - バンディットフィードバック

オンライン凸最適化とリグレット解析

- $X \subseteq \mathbb{R}^d$: 有界閉凸な実行可能領域
- $\mathcal{F} \subseteq \{f: X \rightarrow \mathbb{R}\}$: 連続微分可能かつ有界な凸関数の集合
- $T \in \mathbb{N}$: ラウンド数

- For $t = 1, 2, \dots, T$
 - 学習者（プレイヤー）が解 $x_t \in X$ を決定
 - 環境が目的関数 $f_t \in \mathcal{F}$ を決定
 - 学習者は損失 $f_t(x_t)$ を被り、勾配 $\nabla f_t(x_t)$ を観測

解 x_1 を出力

▶ 勾配 $\nabla f_1(x_1)$ を観測

▶ 解 x_2 を出力

▶ 勾配 $\nabla f_2(x_2)$ を観測

▶ ...

- リグレット R_T : 学習者の意思決定の「良さ」の指標

$$R_T := \sum_{t=1}^T f_t(x_t) - \min_{x^* \in X} \sum_{t=1}^T f_t(x^*)$$

学習者の累積損失

全ての $\{f_t\}_{t=1}^T$ を見たあとでわかる
最適解 x^* が達成する累積損失

オンライン凸最適化問題: 応用例

オンライン線形回帰: [Vov01]

- $t = 1, 2, \dots, T$: 日付をあらわすインデックス
- 説明変数 $\phi_t = \begin{bmatrix} \text{気温}_t \\ \text{湿度}_t \end{bmatrix} \in \mathbb{R}^2$ を使って $y_t = [\text{アイスの販売数}_t] \in \mathbb{R}$ を予測する
- 重み係数パラメタ $w_t = \begin{bmatrix} \beta_{t0} \\ \beta_{t1} \\ \beta_{t2} \end{bmatrix} \in \mathbb{R}^3$ で定まる線形回帰モデル:
 $\hat{y}_t = \beta_{t0} + \beta_{t1} \cdot [\text{気温}_t] + \beta_{t2} \cdot [\text{湿度}_t] = \langle w_t, \phi'_t \rangle$ を使って予測 $\left(\phi'_t = \begin{bmatrix} 1 \\ \phi_t \end{bmatrix} \right)$
- t 日目の予測値 \hat{y}_t を出力したあと, 実際の販売数 y_t を確認
- 予測精度は $f_t(w_t) = (\hat{y}_t - y_t)^2 = (\langle w_t, \phi'_t \rangle - y_t)^2$ で評価

ϕ_1 をもとに \hat{y}_1 を出力
(パラメタ w_1 に依存)

y_1 を観測
(f_1 を観測)

ϕ_2 をもとに \hat{y}_2 を出力
(パラメタ w_2 に依存)

y_2 を観測
(f_2 を観測)

...

オンライン凸最適化問題 応用例

オンライン線形回帰: [Vov01]

- 重み係数パラメタ $w_t \in \mathbb{R}^d$ で定まる線形回帰モデル: $\hat{y}_t = \langle w_t, \phi_t' \rangle$ で予測
- t 日目の予測値 \hat{y}_t を出力したあと, 実際の販売数 y_t を確認
- 予測精度は二乗損失: $f_t(w_t) = (\hat{y}_t - y_t)^2 = (\langle w_t, \phi_t' \rangle - y_t)^2$ で評価

- リグレットの解釈:
$$R_T = \sum_{t=1}^T f_t(w_t) - \min_{x^* \in X} \sum_{t=1}^T f_t(w^*)$$

アルゴリズム
の累積損失

最適な重み係数パラメタ w^*
を使った場合の累積損失

- $f_t(w)$ は凸関数であるから, オンライン凸最適化の例になっている
- 線形回帰モデルに限らず, 損失がモデルパラメタ w の凸関数ならば
オンライン凸最適化の特殊ケースとしてモデル化可能

オンライン凸最適化問題 応用例

オンラインポートフォリオ選択: ^{[Cov91],}
_[HAK06]

- $t = 1, 2, \dots, T$: 月をあらわすインデックス
- $i = 1, 2, \dots, d$: 投資対象（株式銘柄など）をあらわすインデックス
- 第 t 月のはじめに投資比率（株式保有割合） $p_t = \begin{bmatrix} p_{t1} \\ \vdots \\ p_{td} \end{bmatrix} \in \Delta^d$ を決定
- 第 t 月のおわりに利益率（価格倍率） $r_t = \begin{bmatrix} r_{t1} \\ \vdots \\ r_{td} \end{bmatrix} \in \mathbb{R}_{>0}^d$ を観測
- $r_{ti} = \frac{\text{収入金額}}{\text{投資金額}} = \frac{\text{月末の株価}}{\text{月初の株価}}$
- 第 t 月中に、資産総額は $\sum_{i=1}^d p_{ti} r_{ti}$ 倍 = $\langle r_t, p_t \rangle$ 倍に

p_1 を決定

r_1 を観測

p_2 を出力

r_2 を観測

...

オンライン凸最適化問題 応用例

オンラインポートフォリオ選択: [Cov91], [HAK06]

- 第 t 月のはじめに投資比率 $p_t \in \Delta^d$ を, おわりに利益率 $r_t \in \mathbb{R}_{>0}^d$ を観測
- 第 t 月中に, 資産総額は $\sum_{i=1}^N p_{ti} r_{ti}$ 倍 = $\langle r_t, p_t \rangle$ 倍になる
- 初期の資産総額を 1 とすると, 第 T 月のおわりの資産総額は:
 $1 \cdot \langle r_1, p_1 \rangle \cdot \langle r_2, p_2 \rangle \cdots \langle r_T, p_T \rangle = \prod_{t=1}^T \langle r_t, p_t \rangle = \exp(\sum_{t=1}^T \log(\langle r_t, p_t \rangle))$
- $f_t(p_t) = -\log \langle r_t, p_t \rangle$ で目的関数 f_t を定義すると, f_t は凸関数
- リグレットの解釈:

$$R_T = \sum_{t=1}^T f_t(p_t) - \min_{p^* \in \Delta^d} \sum_{t=1}^T f_t(p^*) = \max_{p^* \in \Delta^d} \sum_{t=1}^T \log \langle r_t, p^* \rangle - \sum_{t=1}^T \log \langle r_t, p_t \rangle$$
$$= \max_{p^* \in \Delta^d} \log \frac{\prod_{t=1}^T \langle r_t, p^* \rangle}{\prod_{t=1}^T \langle r_t, p_t \rangle} = \log \frac{\text{最適な投資比率 } p^* \text{ を使った場合の資産総額}}{\text{アルゴリズムが達成した資産総額}}$$

p_1 を決定

r_1 を観測

p_2 を出力

r_2 を観測

...

オンライン凸最適化とリグレット解析

- $X \subseteq \mathbb{R}^d$: 有界閉凸な実行可能領域
- $\mathcal{F} \subseteq \{f: X \rightarrow \mathbb{R}\}$: 連続微分可能かつ有界な凸関数の集合
- $T \in \mathbb{N}$: ラウンド数

- For $t = 1, 2, \dots, T$
 - 学習者（プレイヤー）が解 $x_t \in X$ を決定
 - 環境が目的関数 $f_t \in \mathcal{F}$ を決定
 - 学習者は損失 $f_t(x_t)$ を被り、勾配 $\nabla f_t(x_t)$ を観測

解 x_1 を出力

▶ 勾配 $\nabla f_1(x_1)$ を観測

▶ 解 x_2 を出力

▶ 勾配 $\nabla f_2(x_2)$ を観測

▶ ...

- リグレット R_T : 学習者の意思決定の「良さ」の指標

$$R_T := \sum_{t=1}^T f_t(x_t) - \min_{x^* \in X} \sum_{t=1}^T f_t(x^*)$$

学習者の累積損失

全ての $\{f_t\}_{t=1}^T$ を見たあとでわかる
最適解 x^* が達成する累積損失

一般化：オンライン学習とリグレット解析

- X : 実行可能領域
- $\mathcal{F} \subseteq \{f: X \rightarrow \mathbb{R}\}$: 目的関数（損失関数）のクラス
- $T \in \mathbb{N}$: ラウンド数

- For $t = 1, 2, \dots, T$
 - 学習者（プレイヤー）が解 $x_t \in X$ を決定
 - 環境が目的関数 $f_t \in \mathcal{F}$ を決定
 - 学習者は損失 $f_t(x_t)$ を被り, f_t についてのフィードバック o_t を観測

解 x_1 を出力

o_1 を観測

解 x_2 を出力

o_2 を観測

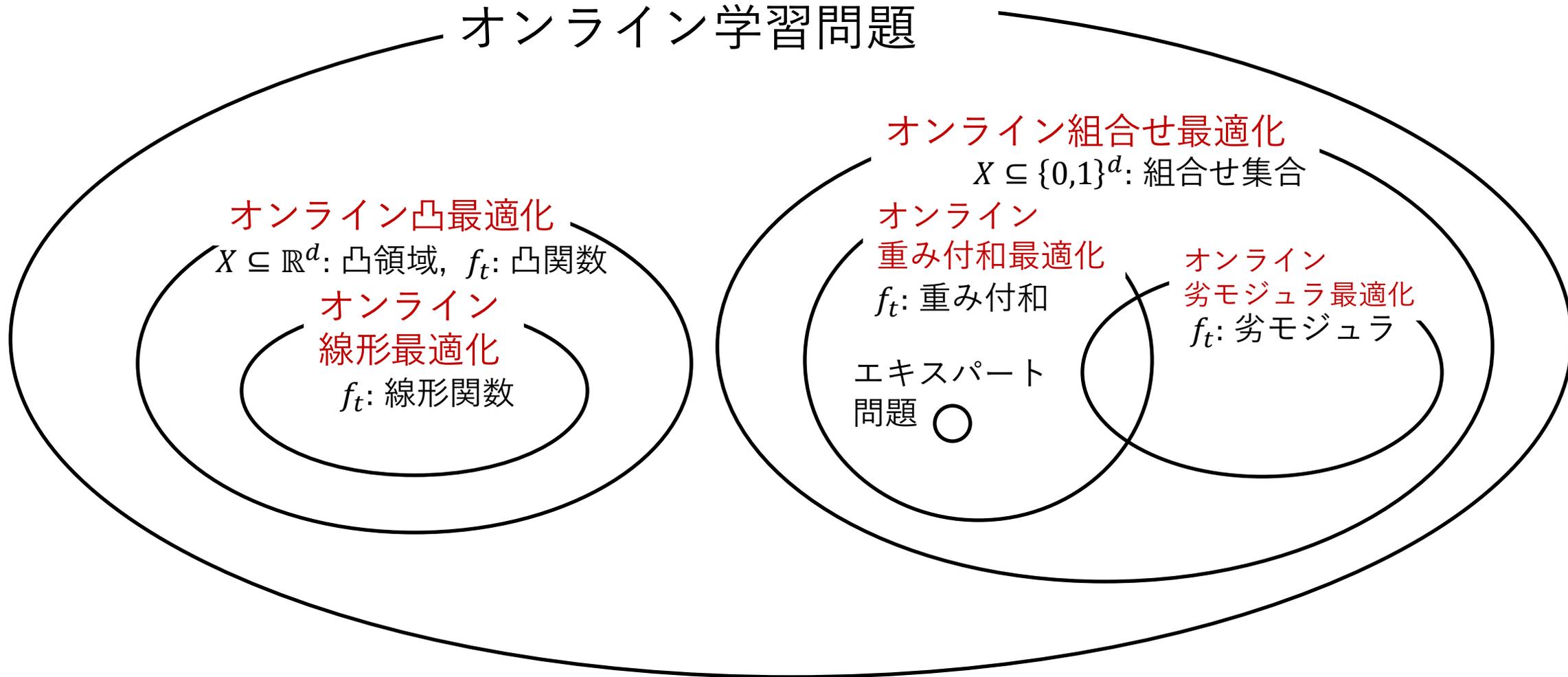
...

- リグレット R_T : 学習者の意思決定の「良さ」の指標

$$R_T := \sum_{t=1}^T f_t(x_t) - \min_{x^* \in X} \sum_{t=1}^T f_t(x^*)$$

問題のクラス

オンライン学習問題



実行可能領域 X , 目的関数 f_t のクラス \mathcal{F} で分類

オンライン組合せ最適化 応用例

オンライン最短経路問題: [AK04], [AK08]

- 事前に重み無し有向グラフ $G = (V, E)$ と始点 $s \in V$, 終点 $g \in V$ が与えられる
- 各ラウンド $t = 1, 2, \dots, T$ において,
経路 P_t を決定したあとで, 枝重み (通行所要時間など) $w_t: E \rightarrow \mathbb{R}_{\geq 0}$ を観測

経路 $P_1 \subseteq E$ を出力

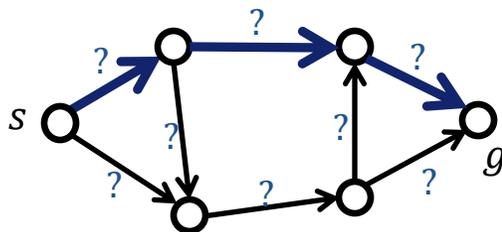
重み $w_1 \in \mathbb{R}_{\geq 0}^E$ を観測

経路 $P_2 \subseteq E$ を出力

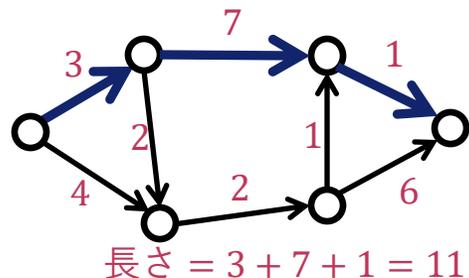
重み $w_2 \in \mathbb{R}_{\geq 0}^E$ を観測

...

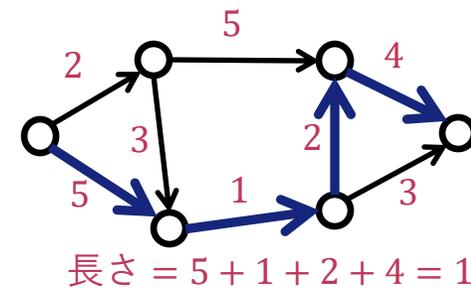
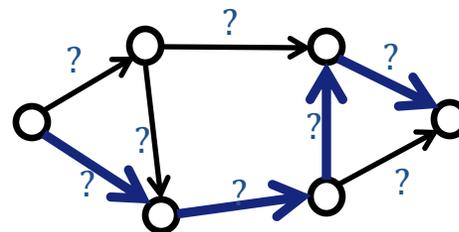
枝重みがわからない状況
で経路をきめる



あとになって、各枝の重み
がわかる



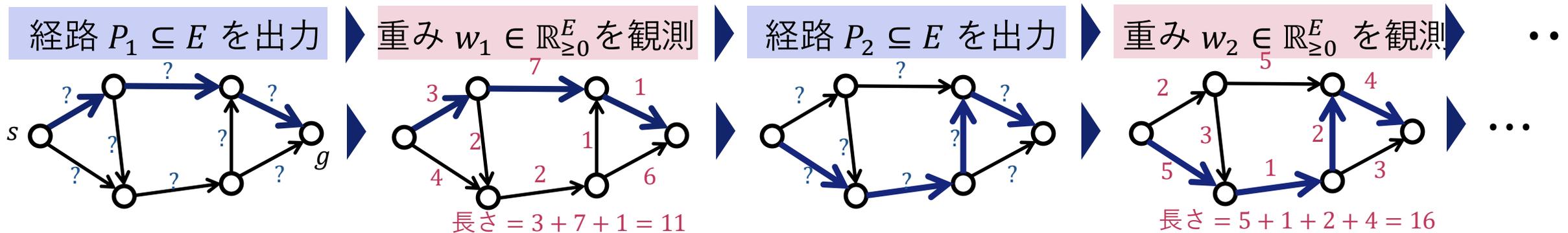
同様に繰り返し
(枝重みは毎回変化)



...

オンライン組合せ最適化 応用例

オンライン最短経路問題: [AK04], [AK08]



- $X \subseteq \{0,1\}^E$: s - g 経路をなす枝集合の指示ベクトル全体
- $x_t = [x_{te}]_{e \in E} \in X$ を P_t の指示ベクトルとする: $x_{te} = 1 (e \in P_t), x_{te} = 0 (e \notin P_t)$
- 第 t ラウンドの経路長は $\sum_{e \in P_t} w_{te} = \langle w_t, x_t \rangle =: f_t(x_t)$ であらわせる
- リグレットの解釈: $R_T = \sum_{t=1}^T f_t(x_t) - \min_{x^* \in X} \sum_{t=1}^T f_t(x^*)$
 - 選んだ経路の長さの総和
 - 最適な経路 $P^* \subseteq E$ を使い続けた場合の経路長の総和
- 同様に、最小コストフロー・最大重みマッチング・ナップサック問題なども、目的関数を適当な $X \subseteq \{0,1\}^d$ 上の重み付和としてあらわし、対応するオンライン最適化を考えることができる

環境モデル／敵対者モデル

環境モデル：目的関数列 $\{f_t\}_t^T \subseteq \mathcal{F}$ の生成プロセス

- 易 \uparrow
- (1) 確率的環境モデル： \mathcal{F} 上の確率分布 D が存在し、 $f_t \sim D$ (i.i.d.)
リグレット解析では（本講義では）期待値を評価
 - (2) 非適応的敵対者モデル： $\{f_t\}_{t=1}^T$ は敵対者によって任意に、事前に決定される
 - (3) 弱適応的敵対者モデル：敵対者が $\{x_1, x_2, \dots, x_{t-1}\}$ を見てから f_t を決定
 - (4) 強適応的敵対者モデル：敵対者が $\{x_1, x_2, \dots, x_t\}$ を見てから f_t を決定
- 難 \downarrow

「敵対者」は最悪ケース解析，つまり学習者にとって最も都合の悪いインスタンス $\{f_t\}$ に対応

$$\sup_{\{f_t\}} \{R_T\} = \sup_{\{f_t\}} \left\{ \sum_{t=1}^T f_t(x_t) - \min_{x^* \in X} \sum_{t=1}^T f_t(x^*) \right\} \leq ??$$

学習者が決定的アルゴリズムを用いるとき，(2) (3) (4) のモデルは等価
(敵対者は $\{x_1, \dots, x_{t-1}\}$ と $\{f_1, \dots, f_{t-1}\}$ から x_t を計算できるため.)

環境モデル／敵対者モデル

環境モデル：目的関数列 $\{f_t\}_t^T \subseteq \mathcal{F}$ の生成プロセス

- 易 \uparrow
- (1) 確率的環境モデル： \mathcal{F} 上の確率分布 D が存在し、 $f_t \sim D$ (i.i.d.)
リグレット解析では（本講義では）期待値を評価
 - (2) 非適応的敵対者モデル： $\{f_t\}_{t=1}^T$ は敵対者によって任意に、事前に決定される
 - (3) **弱適応的敵対者モデル**：敵対者が $\{x_1, x_2, \dots, x_{t-1}\}$ を見てから f_t を決定
 - (4) 強適応的敵対者モデル：敵対者が $\{x_1, x_2, \dots, x_t\}$ を見てから f_t を決定
- 難 \downarrow

「敵対者」は最悪ケース解析，つまり学習者にとって最も都合の悪いインスタンス $\{f_t\}$ に対応

$$\sup_{\{f_t\}} \{R_T\} = \sup_{\{f_t\}} \left\{ \sum_{t=1}^T f_t(x_t) - \min_{x^* \in X} \sum_{t=1}^T f_t(x^*) \right\} \leq ??$$

学習者が決定的アルゴリズムを用いるとき，(2) (3) (4) のモデルは等価
(敵対者は $\{x_1, \dots, x_{t-1}\}$ と $\{f_1, \dots, f_{t-1}\}$ から x_t を計算できるため.)

この講義では，断らない限り **(3) 弱適応的敵対者モデルを想定**

オンライン勾配降下法

仮定

- $X \subseteq \mathbb{R}^d$: 有界閉凸, $\max_{x,y \in X} \|x - y\|_2 \leq D$
- $f_t: X \rightarrow \mathbb{R}$: 凸かつ連続微分可能, $\forall t, \forall x \in X, \|\nabla f_t(x)\|_2 \leq G$

オンライン勾配降下法

- 入力 $x_1 \in X$: 任意の初期解, $\eta > 0$: ステップサイズ
- For $t = 1, 2, \dots$
 - x_t を出力し, 勾配 $g_t := \nabla f_t(x_t)$ を観測
 - $\tilde{x}_{t+1} = x_t - \eta g_t$
 - $x_{t+1} = \text{Proj}_X(\tilde{x}_{t+1}) \in \arg \min_{x \in X} \|x - \tilde{x}_{t+1}\|_2$

定理
$$R_T := \sum_{t=1}^T f_t(x_t) - \min_{x^* \in X} \sum_{t=1}^T f_t(x^*) \leq \frac{\eta}{2} G^2 T + \frac{D^2}{2\eta}$$

したがって $\eta = \frac{D}{G\sqrt{T}}$ とおくと $R_T \leq DG\sqrt{T}$

オンライン勾配降下法

仮定: $f_t: X \rightarrow \mathbb{R}$: 凸関数, $\max_{x, y \in X} \|x - y\|_2 \leq D$, $\|\nabla f_t(x)\|_2 \leq G$

[Step 1] $\tilde{x}_{t+1} = x_t - \eta g_t$ ($g_t := \nabla f_t(x_t)$) [Step 2] $x_{t+1} = \text{Proj}_X(\tilde{x}_{t+1}) \in \arg \min_{x \in X} \|x - \tilde{x}_{t+1}\|_2$

定理 $R_T := \sum_{t=1}^T f_t(x_t) - \min_{x^* \in X} \sum_{t=1}^T f_t(x^*) \leq \frac{\eta}{2} G^2 T + \frac{D^2}{2\eta}$

(証明)

- f_t の凸性より、 $R_T = \sum_{t=1}^T (f_t(x_t) - f_t(x^*)) \leq \sum_{t=1}^T \langle g_t, x_t - x^* \rangle$
- [Step 1] より、 $\tilde{x}_{t+1} - x^* = x_t - x^* - \eta g_t$ であるから、
 $\|\tilde{x}_{t+1} - x^*\|_2^2 = \|x_t - x^* - \eta g_t\|_2^2 = \|x_t - x^*\|_2^2 + \|\eta g_t\|_2^2 - 2\eta \langle g_t, x_t - x^* \rangle$
- 変形して、 $\langle g_t, x_t - x^* \rangle = \frac{1}{2\eta} (\|x_t - x^*\|_2^2 - \|\tilde{x}_{t+1} - x^*\|_2^2) + \frac{\eta}{2} \|g_t\|_2^2$
- [Step 2] と $x^* \in X$ より角 $x^* x_{t+1} \tilde{x}_{t+1}$ は鈍角だから $\|\tilde{x}_{t+1} - x^*\|_2^2 \geq \|x_{t+1} - x^*\|_2^2$
- まとめると、 $R_T \leq \sum_{t=1}^T \left(\frac{1}{2\eta} (\|x_t - x^*\|_2^2 - \|x_{t+1} - x^*\|_2^2) + \frac{\eta}{2} \|g_t\|_2^2 \right)$
 $\leq \frac{1}{2\eta} \|x_1 - x^*\|_2^2 + \frac{\eta}{2} \sum_{t=1}^T \|g_t\|_2^2 \leq \frac{1}{2\eta} D^2 + \frac{\eta}{2} L^2 T$

リグレット上界の応用例 (1) Online-to-batch 変換

機械学習における典型的な最適化問題

確率変数 Z を用いて $\bar{f}(x) := \mathbf{E}_Z[f(x; Z)]$ の形式で表せる目的関数

Z の確率分布 D の厳密な情報は未知, D からの独立なサンプル $z_1, z_2, \dots, z_T \sim D$ にアクセス可能

例: 線形回帰問題

分布 D に従うデータ点 $z_t = (a_t, b_t) \in \mathbb{R}^d \times \mathbb{R}$ が与えられたとき, 回帰係数 $x \in \mathbb{R}^d$ を決めて, $\bar{f}(x) := \mathbf{E}_{Z \sim D}[f(x; Z)] = \mathbf{E}_{(a,b) \sim D}[(a^\top x - b)^2]$ を小さくしたい. ($f(x; a, b) := (a^\top x - b)^2$)

定理 $f_t(x) := f(x; z_t)$ で与えられるオンライン凸最適化問題において, $\bar{x}_T = \frac{1}{T} \sum_{t=1}^T x_t$ とすると

$$\mathbf{E}_{z_1, \dots, z_T} [\bar{f}(\bar{x}_T)] - \min_{x^* \in X} \bar{f}(x^*) \leq \frac{1}{T} \mathbf{E}_{z_1, \dots, z_T} [R_T]$$

(証明)

- 任意の t と $x^* \in X$ に対して, $\mathbf{E}[f_t(x^*)] = \mathbf{E}[f(x^*; z_t)] = \bar{f}(x^*)$
- z_t は $\{z_1, \dots, z_{t-1}\}$ と独立であるから, z_t と x_t も独立なので $\mathbf{E}[f_t(x_t)] = \mathbf{E}[f(x_t; z_t)] = \mathbf{E}[\bar{f}(x_t)]$
- $\frac{1}{T} E[R_T] \geq \frac{1}{T} E[\sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T f_t(x^*)] = E\left[\frac{1}{T} \sum_{t=1}^T \bar{f}(x_t)\right] - \bar{f}(x^*) \geq E\left[\bar{f}\left(\frac{1}{T} \sum_{t=1}^T x_t\right)\right] - \bar{f}(x^*)$

リグレット上界の応用例 (1) Online-to-batch 変換

機械学習における典型的な最適化問題

確率変数 Z を用いて $\bar{f}(x) := \mathbf{E}_Z[f(x; Z)]$ の形式で表せる目的関数

Z の確率分布 D の厳密な情報は未知, D からの独立なサンプル $z_1, z_2, \dots, z_T \sim D$ にアクセス可能

例: 線形回帰問題

分布 D に従うデータ点 $z_t = (a_t, b_t) \in \mathbb{R}^d \times \mathbb{R}$ が与えられたとき, 回帰係数 $x \in \mathbb{R}^d$ を決めて, $\bar{f}(x) := \mathbf{E}_{Z \sim D}[f(x; Z)] = \mathbf{E}_{(a,b) \sim D}[(a^\top x - b)^2]$ を小さくしたい. ($f(x; a, b) := (a^\top x - b)^2$)

定理 $f_t(x) := f(x; z_t)$ で与えられるオンライン凸最適化問題において, $\bar{x}_T = \frac{1}{T} \sum_{t=1}^T x_t$ とすると

$$\mathbf{E}_{z_1, \dots, z_T} [\bar{f}(\bar{x}_T)] - \min_{x^* \in X} \bar{f}(x^*) \leq \frac{1}{T} \mathbf{E}_{z_1, \dots, z_T} [R_T]$$

(解釈)

- $R_T = O(C\sqrt{T})$ ならば, $[\text{RHS}] = O\left(\frac{C}{\sqrt{T}}\right)$ より, $T = O\left(\frac{C^2}{\epsilon^2}\right)$ で $[\text{RHS}] \leq \epsilon$ にできる.
- オンラインで劣線形リグレット: $R_T = o(T)$ を達成することは, 少なくとも, \bar{f} を最小化するのと同程度以上の難しさ

リグレット上界の応用例（２） ミニマックス定理

関数 $F: X \times Y \rightarrow \mathbb{R}$ を凸-凹関数 とする：

- X は凸集合で、任意の $y \in Y$ に対して $x \mapsto F(x, y)$ が凸関数
- Y は凸集合で、任意の $x \in X$ に対して $y \mapsto F(x, y)$ が凹関数

$f_t(x) := F(x, y_t)$ で定まる X 上オンライン凸最適化のリグレット： $R_T = \sum_{t=1}^T f_t(x_t) - \min_{x^* \in X} \sum_{t=1}^T f_t(x^*)$

$f'_t(y) := -F(x_t, y)$ で定まる Y 上オンライン凸最適化のリグレット： $R'_T = \sum_{t=1}^T f'_t(y_t) - \min_{y^* \in Y} \sum_{t=1}^T f'_t(y^*)$

定理 上記の設定で $\bar{x}_T = \frac{1}{T} \sum_{t=1}^T x_t$, $\bar{y}_T = \frac{1}{T} \sum_{t=1}^T y_t$ とすると、任意の $x \in X, y \in Y$ に対し

$$\max_{y \in Y} F(\bar{x}_T, y) - \min_{x \in X} F(x, \bar{y}_T) \leq \frac{R_T + R'_T}{T}$$

(証明)

- 任意の $x \in X, y \in Y$ に対し,

$$\begin{aligned} R_T + R'_T &\geq (\sum_{t=1}^T F(x_t, y_t) - \sum_{t=1}^T F(x, y_t)) + (-\sum_{t=1}^T F(x_t, y_t) + \sum_{t=1}^T F(x_t, y)) \\ &= \sum_{t=1}^T F(x_t, y) - \sum_{t=1}^T F(x, y_t) \geq T(F(\bar{x}_T, y) - F(x, \bar{y}_T)) \end{aligned}$$

リグレット上界の応用例（２） ミニマックス定理

関数 $F: X \times Y \rightarrow \mathbb{R}$ を凸-凹関数 とする：

- X は凸集合で、任意の $y \in Y$ に対して $x \mapsto F(x, y)$ が凸関数
- Y は凸集合で、任意の $x \in X$ に対して $y \mapsto F(x, y)$ が凹関数

$f_t(x) := F(x, y_t)$ で定まる X 上オンライン凸最適化のリグレット： $R_T = \sum_{t=1}^T f_t(x_t) - \min_{x^* \in X} \sum_{t=1}^T f_t(x^*)$

$f'_t(y) := -F(x_t, y)$ で定まる Y 上オンライン凸最適化のリグレット： $R'_T = \sum_{t=1}^T f'_t(y_t) - \min_{y^* \in Y} \sum_{t=1}^T f'_t(y^*)$

定理 上記の設定で $\bar{x}_T = \frac{1}{T} \sum_{t=1}^T x_t$, $\bar{y}_T = \frac{1}{T} \sum_{t=1}^T y_t$ とすると、任意の $x \in X, y \in Y$ に対し

$$\max_{y \in Y} F(\bar{x}_T, y) - \min_{x \in X} F(x, \bar{y}_T) \leq \frac{R_T + R'_T}{T}$$

系 X, Y ：有界閉凸, $F: X \times Y \rightarrow \mathbb{R}$ ：凸-凹関数, $F: x, y$ それぞれについて連続微分可能,

$\{\nabla_x F(x, y) \mid x \in X, y \in Y\}, \{\nabla_y F(x, y) \mid x \in X, y \in Y\}$ ：有界

を仮定すると, $\min_{x \in X} \max_{y \in Y} F(x, y) = \max_{y \in Y} \min_{x \in X} F(x, y)$

(証明) $[\geq]$ は任意の F で成立. $[\leq]$ は, $R_T = o(T)$ なるアルゴリズムの存在と定理より従う

リグレット上界の応用例 (2) ミニマックス定理：補足①

$f_t(x) := F(x, y_t)$ で定まる X 上オンライン凸最適化のリグレット： $R_T = \sum_{t=1}^T f_t(x_t) - \min_{x^* \in X} \sum_{t=1}^T f_t(x^*)$

$f'_t(y) := -F(x_t, y)$ で定まる Y 上オンライン凸最適化のリグレット： $R'_T = \sum_{t=1}^T f'_t(y_t) - \min_{y^* \in Y} \sum_{t=1}^T f'_t(y^*)$

定理 上記の設定で $\bar{x}_T = \frac{1}{T} \sum_{t=1}^T x_t$, $\bar{y}_T = \frac{1}{T} \sum_{t=1}^T y_t$ とすると, 任意の $x \in X, y \in Y$ に対し

$$\max_{y \in Y} F(\bar{x}_T, y) - \min_{x \in X} F(x, \bar{y}_T) \leq \frac{R_T + R'_T}{T}$$

- これと $\min_{x \in X} \max_{y \in Y} F(x, y) = \max_{y \in Y} \min_{x \in X} F(x, y)$ を認めると,

$$\left(\max_{y \in Y} F(\bar{x}_T, y) - \min_{x \in X} \max_{y \in Y} F(x, y) \right) + \left(\max_{y \in Y} \min_{x \in X} F(x, y) - \min_{x \in X} F(x, \bar{y}_T) \right) \leq \frac{R_T + R'_T}{T}$$

つまり, $G(x) := \max_{y \in Y} F(x, y)$, $H(y) := \min_{x \in X} F(x, y)$ に対し

$$\left(G(\bar{x}_T) - \min_{x \in X} G(x) \right) + \left(\max_{y \in Y} H(y) - H(\bar{y}_T) \right) \leq \frac{R_T + R'_T}{T}$$

()内はいずれも非負だから, どちらも $\leq \frac{R_T + R'_T}{T}$.

リグレット上界の応用例（２） ミニマックス定理：補足②

$f_t(x) := F(x, y_t)$ で定まる X 上オンライン凸最適化のリグレット： $R_T = \sum_{t=1}^T f_t(x_t) - \min_{x^* \in X} \sum_{t=1}^T f_t(x^*)$

$f'_t(y) := -F(x_t, y)$ で定まる Y 上オンライン凸最適化のリグレット： $R'_T = \sum_{t=1}^T f'_t(y_t) - \min_{y^* \in Y} \sum_{t=1}^T f'_t(y^*)$

定理 上記の設定で $\bar{x}_T = \frac{1}{T} \sum_{t=1}^T x_t$, $\bar{y}_T = \frac{1}{T} \sum_{t=1}^T y_t$ とすると、任意の $x \in X, y \in Y$ に対し

$$\max_{y \in Y} F(\bar{x}_T, y) - \min_{x \in X} F(x, \bar{y}_T) \leq \frac{R_T + R'_T}{T}$$

- 上記の定理を示す上では、必ずしも x_t, y_t がオンライン凸最適化のプロトコルに従っている必要はない
- 例えば、 $x_t \in \operatorname{argmin}_{x \in X} f_t(x) = \operatorname{argmin}_{x \in X} F(x, y_t)$ のように、 x_t が f_t に依存していたとしても、定理の不等式は依然として成立
- この場合、明らかに $R_T \leq 0$ だから $\max_{y \in Y} F(\bar{x}_T, y) - \min_{x \in X} F(x, \bar{y}_T) \leq \frac{R'_T}{T}$

リグレット上界の応用例（２） ミニマックス定理：補足③

$f_t(x) := F(x, y_t)$ で定まる X 上オンライン凸最適化のリグレット： $R_T = \sum_{t=1}^T f_t(x_t) - \min_{x^* \in X} \sum_{t=1}^T f_t(x^*)$

$f'_t(y) := -F(x_t, y)$ で定まる Y 上オンライン凸最適化のリグレット： $R'_T = \sum_{t=1}^T f'_t(y_t) - \min_{y^* \in Y} \sum_{t=1}^T f'_t(y^*)$

定理 上記の設定で $\bar{x}_T = \frac{1}{T} \sum_{t=1}^T x_t$, $\bar{y}_T = \frac{1}{T} \sum_{t=1}^T y_t$ とすると, 任意の $x \in X, y \in Y$ に対し

$$\max_{y \in Y} F(\bar{x}_T, y) - \min_{x \in X} F(x, \bar{y}_T) \leq \frac{R_T + R'_T}{T}$$

系 X, Y : 有界閉凸, $F: X \times Y \rightarrow \mathbb{R}$: 凸-凹関数, $F: x, y$ それぞれについて連続微分可能,
 $\{\nabla_x F(x, y) \mid x \in X, y \in Y\}, \{\nabla_y F(x, y) \mid x \in X, y \in Y\}$: 有界
を仮定すると, $\min_{x \in X} \max_{y \in Y} F(x, y) = \max_{y \in Y} \min_{x \in X} F(x, y)$

- 系の等式が成り立つための十分条件は大幅に緩めることができる (**Sion のミニマックス定理**)
 - X, Y の片方は非有界でもいい
 - 連続微分可能や勾配の有界性の仮定は Upper/lower semicontinuous に緩和できる
 - 凸-凹性は quasi-凸-凹性 に緩和できる

エキスパート問題

- $A := [N] := \{1, 2, \dots, N\}$: 有限個の選択肢の集合
- $\ell_t = (\ell_t(i))_{i \in [N]} \in [0, 1]^N$: 損失ベクトル
- $T \in \mathbb{N}$: ラウンド数

- For $t = 1, 2, \dots, T$
 - 環境が損失ベクトル $\ell_t \in [0, 1]^N$ を決定
 - 学習者（プレイヤー）が $i_t \in [N]$ を決定
 - 学習者は損失 $\ell_t(i_t)$ を被り，損失ベクトル ℓ_t を観測

- 弱適応的敵対者モデルを想定： ℓ_t は $\{i_1, i_2, \dots, i_{t-1}\}$ に依存してもよい
- リグレット R_T : 学習者の意思決定の「良さ」の指標

$$R_T := \max_{i^* \in [N]} \mathbf{E} \left[\sum_{t=1}^T \ell_t(i_t) - \sum_{t=1}^T \ell_t(i^*) \right]$$

ここで $\mathbf{E}[\cdot]$ は学習者および環境のすべてのランダム性についての期待値

エキスパート問題：オンライン凸最適化への帰着

- $A := [N] := \{1, 2, \dots, N\}$
- $\ell_t = (\ell_t(i))_{i \in [N]} \in [0, 1]^N$
- $T \in \mathbb{N}$: ラウンド数
- $X = \Delta(N) := \{x \in \mathbb{R}_{\geq 0}^N \mid \|x\|_1 = 1\}$: 確率単体
- $f_t: X \rightarrow [0, 1]$, $f_t(x) = \ell_t^\top x = \sum_{i=1}^N x(i)\ell_t(i)$
($\nabla f_t(x) = \ell_t$)

- For $t = 1, 2, \dots, T$
 - 環境が損失ベクトル $\ell_t \in [0, 1]^N$ を決定
 - 学習者（プレイヤー）は分布 $x_t \in \Delta(N)$ を決定し, $i_t \sim x_t$ を選択: $\Pr(i_t = i) = x_t(i)$
 - 学習者は損失 $\ell_t(i_t)$ を被り, 損失ベクトル ℓ_t を観測 $\implies \mathbf{E}[\ell_t(i_t)] = \ell_t^\top x_t$

- 弱適応的敵対者モデルを想定: ℓ_t は $\{i_1, i_2, \dots, i_{t-1}\}$ に依存してもよい
- リグレット R_T : 学習者の意思決定の「良さ」の指標

$$R_T := \max_{i^* \in [N]} \mathbf{E} \left[\sum_{t=1}^T \ell_t(i_t) - \sum_{t=1}^T \ell_t(i^*) \right] = \max_{x^* \in \Delta(N)} \mathbf{E} \left[\sum_{t=1}^T \ell_t^\top x_t - \sum_{t=1}^T \ell_t^\top x^* \right]$$

ここで $\mathbf{E}[\cdot]$ は学習者および環境のすべてのランダム性についての期待値

エキスパート問題：オンライン凸最適化としての定式化

仮定

- $X = \Delta(N) := \{x \in \mathbb{R}_{\geq 0}^N \mid \|x\|_1 = 1\}$ (確率単体)
- $f_t: X \rightarrow \mathbb{R} : \exists \ell_t \in [0,1]^N, f_t(x) = \ell_t^\top x \quad (\implies \nabla f_t(x) = \ell_t)$

オンライン勾配降下法を用いた場合：

- $\max_{x,y \in X} \|x - y\|_2 \leq D, \|\nabla f_t(x)\|_2 \leq G$ のもとで, $R_T = O(DG\sqrt{T})$
- $D = O(1), G = O(\sqrt{N})$ より, $R_T = O(\sqrt{NT})$
- これは最適か？

エキスパート問題に対するHedgeアルゴリズム

仮定

- $X = \Delta(N) := \{x \in \mathbb{R}_{\geq 0}^N \mid \|x\|_1 = 1\}$ (確率単体)
- $f_t: X \rightarrow \mathbb{R} : \exists \ell_t \in [0,1]^N, f_t(x) = \ell_t^\top x \quad (\Rightarrow \nabla f_t(x) = \ell_t)$

Hedgeアルゴリズム/乗算型重み更新 (MWU)

- 入力: $\eta > 0$: ステップサイズ
- $w_1 = (w_1(i))_{i \in [N]} \in \mathbb{R}^N$ を $w_1(i) = 1 (\forall i \in [N])$ で初期化
- For $t = 1, 2, \dots$
 - $x_t = \frac{w_t}{\|w_t\|_1}$ (x_t は w_t に比例する確率分布)
 - $w_{t+1}(i) = w_t(i) \exp(-\eta \ell_t(i)) \quad (\forall i \in [N])$

定理 $\sum_{t=1}^T \ell_t^\top x_t - \min_{i^* \in [N]} \sum_{t=1}^T \ell_t(i^*) \leq \frac{1}{\eta} \log N + \frac{\eta}{2} T$

したがって $\eta = \Theta\left(\frac{\sqrt{\log N}}{\sqrt{T}}\right)$ とおくと $R_T = O(\sqrt{T \log N})$

エキスパート問題に対するHedgeアルゴリズム

仮定 $\ell_t = (\ell_t(i))_{i \in [N]} \in [0,1]^N$

$$w_1(i) = 1 (i \in [N]), \quad w_{t+1}(i) = w_t(i) \exp(-\eta \ell_t(i)), \quad x_t = \frac{w_t}{\|w_t\|_1}$$

定理 $\sum_{t=1}^T \ell_t^\top x_t - \min_{i^* \in [N]} \sum_{t=1}^T \ell_t(i^*) \leq \frac{1}{\eta} \log N + \frac{\eta}{2} T$

(証明) $W_t := \sum_{i=1}^N w_t(i) = \|w_t\|_1$ とすると, $x_t = \frac{w_t}{W_t}$, $W_t x_t = w_t$

- $w_t(i) = w_{t-1}(i) \exp(-\eta \ell_{t-1}(i)) = \dots = \exp(-\eta \sum_{s=1}^{t-1} \ell_s(i))$ より, 任意の $i^* \in [N]$ に対して
- $W_{T+1} \geq w_{T+1}(i^*) = \exp(-\eta \sum_{t=1}^T \ell_t(i^*))$ したがって $\log W_{T+1} \geq -\eta \sum_{t=1}^T \ell_t(i^*) \quad \dots (*)$
- $W_{t+1} = \sum_{i=1}^N w_t(i) \exp(-\eta \ell_t(i)) \leq \sum_{i=1}^N w_t(i) \left(1 - \eta \ell_t(i) + \frac{1}{2} \eta^2 \ell_t(i)^2\right) \quad (\because x \leq 0 \Rightarrow \exp(x) \leq 1 + x + \frac{1}{2} x^2)$
 $= W_t \sum_{i=1}^N x_t(i) \left(1 - \eta \ell_t(i) + \frac{1}{2} \eta^2 \ell_t(i)^2\right) = W_t \left(1 - \eta x_t^\top \ell_t + \frac{\eta^2}{2} \sum_{i=1}^N \ell_t(i)^2 x_t(i)\right)$
- $\log W_{T+1} = \log W_1 + \sum_{t=1}^T \log \frac{W_{t+1}}{W_t} \leq \log N + \sum_{t=1}^T \log \left(1 - \eta x_t^\top \ell_t + \frac{\eta^2}{2} \sum_{i=1}^N \ell_t(i)^2 x_t(i)\right)$
 $\leq \log N - \eta \sum_{t=1}^T \left(x_t^\top \ell_t + \frac{\eta}{2} \sum_{i=1}^N \ell_t(i)^2 x_t(i)\right) \quad (\because \log(x+1) \leq x) \quad \text{これと (*) より,}$
- $\sum_{t=1}^T \ell_t^\top x_t - \min_{i^* \in [N]} \sum_{t=1}^T \ell_t(i^*) \leq \frac{1}{\eta} \log N + \frac{\eta}{2} \sum_{t=1}^T \sum_{i=1}^N \ell_t(i)^2 x_t(i) \leq \frac{1}{\eta} \log N + \frac{\eta}{2} T$

リグレット上界の応用例 (2) ミニマックス定理

$f_t(x) := F(x, y_t)$ のオンライン凸最適化のリグレット : $R_T = \sum_{t=1}^T f_t(x_t) - \min_{x^* \in X} \sum_{t=1}^T f_t(x^*)$

$f'_t(y) := -F(x_t, y)$ のオンライン凸最適化のリグレット : $R'_T = \sum_{t=1}^T f'_t(y_t) - \min_{y^* \in Y} \sum_{t=1}^T f'_t(y^*)$

定理 $\bar{x}_T = \frac{1}{T} \sum_{t=1}^T x_t, \bar{y}_T = \frac{1}{T} \sum_{t=1}^T y_t \Rightarrow \forall x \in X, \forall y \in Y, F(\bar{x}_T, y) - F(x, \bar{y}_T) \leq \frac{R_T + R'_T}{T}$

特殊ケース：二人零和ゲーム

- $X = \Delta(m) := \{x \in [0,1]^m \mid x_1 + x_2 + \dots + x_m = 1\}, Y = \Delta(n)$ (確率単体)
 - X, Y : 行プレイヤー, 列プレイヤーの混合戦略の集合
- $F(x, y) = x^\top A y, A \in [-1,1]^{m \times n}$: 利得/損失行列
 - 行プレイヤーが戦略 x を, 列プレイヤーが戦略 y を選んだとき
行プレイヤーの利得は $x^\top A y$, 列プレイヤーの利得は $-x^\top A y$
- 任意の $x \in \Delta(m), y \in \Delta(n)$ に対して

$$F(\bar{x}_T, y) - F(x, \bar{y}_T) = \bar{x}_T^\top A y - x^\top A \bar{y}_T = (\bar{x}_T^\top A y - \bar{x}_T^\top A \bar{y}_T) + (\bar{x}_T^\top A \bar{y}_T - x^\top A \bar{y}_T) \leq \frac{R_T + R'_T}{T} = O\left(\sqrt{\frac{\log(mn)}{T}}\right)$$

$\Rightarrow T = O\left(\frac{\log(mn)}{\epsilon^2}\right)$ で, (\bar{x}_T, \bar{y}_T) は二人零和ゲームの ϵ -近似 Nash 均衡

講義の概要

- リグレット解析の枠組みとその応用
 - オンライン凸最適化問題とオンライン勾配降下法
 - Online-to-batch 変換
 - ミニマックス定理
 - エキスパート問題と Hedge アルゴリズム
- 正則化を用いたアルゴリズムとリグレット解析
 - オンライン鏡像降下法
 - 多腕バンディット問題とEXP3アルゴリズム
- オンライン組合せ最適化
 - 完全情報フィードバック
 - セミバンディットフィードバック
 - バンディットフィードバック

オンライン勾配降下法：別解釈

オンライン勾配降下法

- 入力： $x_1 \in X$ ：任意の初期解， $\eta > 0$ ：ステップサイズ
- For $t = 1, 2, \dots$
 - x_t を出力し，勾配 $g_t := \nabla f_t(x_t)$ を観測
 - $\tilde{x}_{t+1} = x_t - \eta g_t$
 - $x_{t+1} = \text{Proj}_X(\tilde{x}_{t+1}) \in \arg \min_{x \in X} \|x - \tilde{x}_{t+1}\|_2$

- オンライン勾配降下の別表記：

$$x_{t+1} = \arg \min_{x \in X} \left\{ \langle g_t, x \rangle + \frac{1}{2\eta} \|x - x_t\|_2^2 \right\}$$

実際， $\langle g_t, x \rangle + \frac{1}{2\eta} \|x - x_t\|_2^2 = \frac{1}{2\eta} \|x - (x_t - \eta g_t)\|_2^2 - \frac{\eta}{2} \|g_t\|_2^2 = \frac{1}{2\eta} \|x - \tilde{x}_{t+1}\|_2^2 + \text{const}$

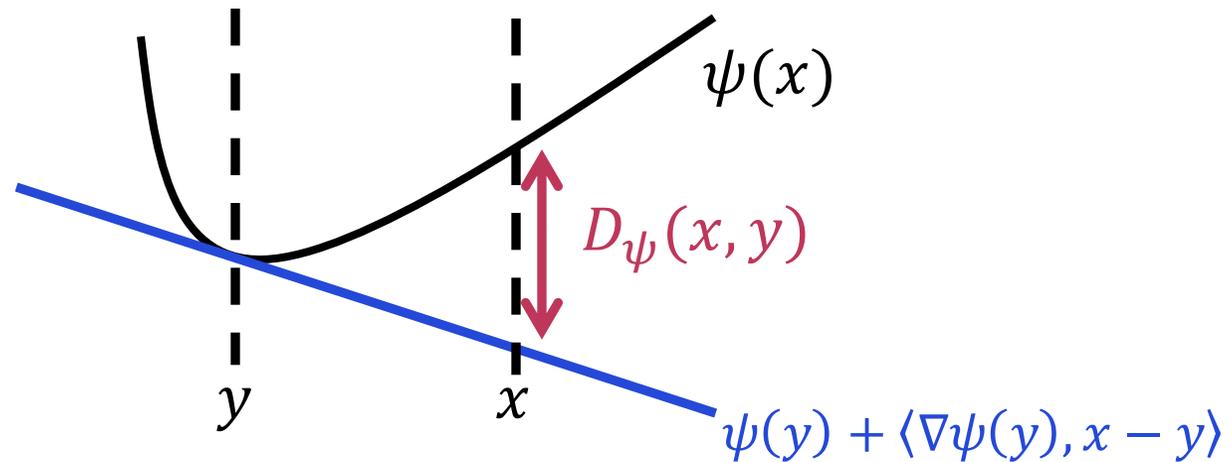
解釈： x_t から離れないようにしつつ， $\langle g_t, x \rangle$ が小さくなるように動かす

- ユークリッドノルム $\|\cdot\|_2$ の代わりに別の「距離」に一般化することを考えてみる

Bregman 距離

- 微分可能な凸関数 $\psi: X \rightarrow \mathbb{R}$ に対し, Bregman 距離 (Bregman divergence) $D_\psi(\cdot, \cdot)$ を次のように定める:

$$D_\psi(x, y) = \psi(x) - (\psi(y) + \langle \nabla \psi(y), x - y \rangle)$$



- 例
 - $\psi(x) = \frac{1}{2} \|x\|_2^2$ のとき, $D_\psi(y, x) = \frac{1}{2} \|x - y\|_2^2$
 - $X = \Delta(N)$ (確率単体), $\psi(x) = \sum_{i=1}^N x_i \log x_i$ (Negentropy) のとき,
 $D_\psi(y, x) = \sum_{i=1}^N y_i \log \frac{y_i}{x_i}$ (KL divergence)
- 注意: Bregman 距離は一般には距離ではない! (対称とは限らない)

オンライン鏡像降下法

$$D_\psi(y, x) = \psi(y) - (\psi(x) + \langle \nabla \psi(x), y - x \rangle)$$

- オンライン勾配降下の別表記：

$$x_{t+1} = \arg \min_{x \in X} \left\{ \langle g_t, x \rangle + \frac{1}{2\eta} \|x - x_t\|_2^2 \right\}$$

- オンライン鏡像降下法： 適当な凸関数 $\psi: X \rightarrow \mathbb{R}$ を用いて

$$x_{t+1} = \arg \min_{x \in X} \left\{ \langle g_t, x \rangle + \frac{1}{\eta} D_\psi(x, x_t) \right\}$$

- 例: $X = \Delta(N)$ (確率単体), $\psi(x) = \sum_{i=1}^N x_i \log x_i$ (Negentropy) のとき Hedge アルゴリズム

(\because) $\langle g_t, x \rangle + \frac{1}{\eta} D_\psi(x, x_t)$ の x についての微分をとると

$$g_t + \frac{1}{\eta} (\nabla \psi(x) - \nabla \psi(x_t)) = \left(g_t(i) + \frac{1}{\eta} (\log x(i) - \log x_t(i)) \right)_{i \in [N]} \quad \text{Lagrange の未定乗数法より,}$$

$$\begin{aligned} g_t(i) + \frac{1}{\eta} (\log x_{t+1}(i) - \log x_t(i)) = \text{const} &\implies \log x_{t+1}(i) = \log x_t(i) - \eta g_t(i) + \text{const} \\ &\implies x_{t+1}(i) \propto x_t(i) \exp(-\eta g_t(i)) \end{aligned}$$

オンライン鏡像降下法 解釈

$$D_\psi(y, x) = \psi(y) - (\psi(x) + \langle \nabla\psi(x), y - x \rangle)$$

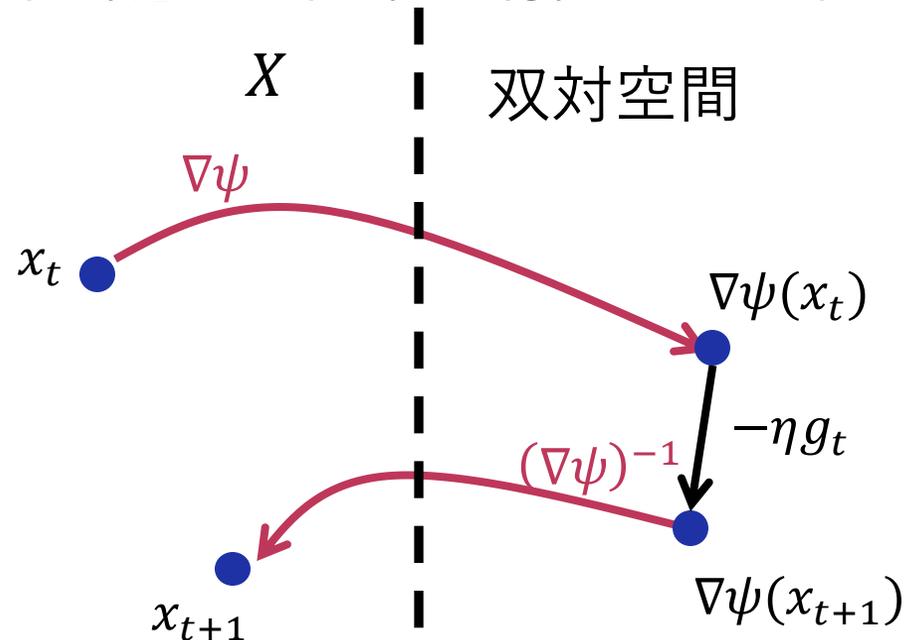
- オンライン鏡像降下法： 適当な凸関数 $\psi: X \rightarrow \mathbb{R}$ を用いて

$$x_{t+1} = \arg \min_{x \in X} \left\{ \langle g_t, x \rangle + \frac{1}{\eta} D_\psi(x, x_t) \right\}$$

- 簡単のため制約条件がないケースを考えると,

$$g_t + \frac{1}{\eta} (\nabla\psi(x_{t+1}) - \nabla\psi(x_t)) = 0$$

$$\Leftrightarrow \nabla\psi(x_{t+1}) = \nabla\psi(x_t) - \eta g_t \quad \text{写像 } \nabla\psi \text{ で写した世界で勾配降下?}$$



オンライン鏡像降下法 解釈

$$D_\psi(y, x) = \psi(y) - (\psi(x) + \langle \nabla\psi(x), y - x \rangle)$$

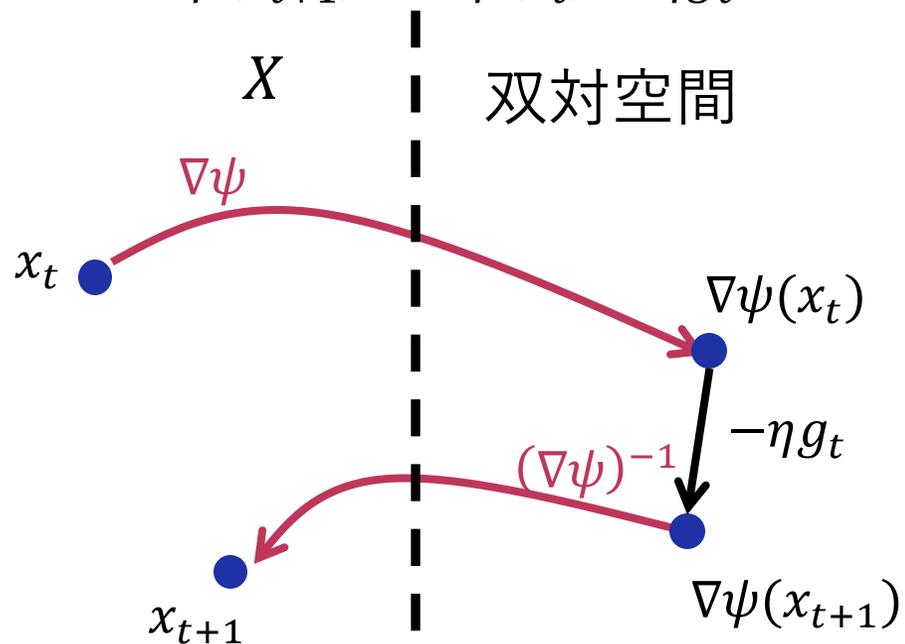
- オンライン鏡像降下法： 適当な凸関数 $\psi: X \rightarrow \mathbb{R}$ を用いて

$$x_{t+1} = \arg \min_{x \in X} \left\{ \langle g_t, x \rangle + \frac{1}{\eta} D_\psi(x, x_t) \right\}$$

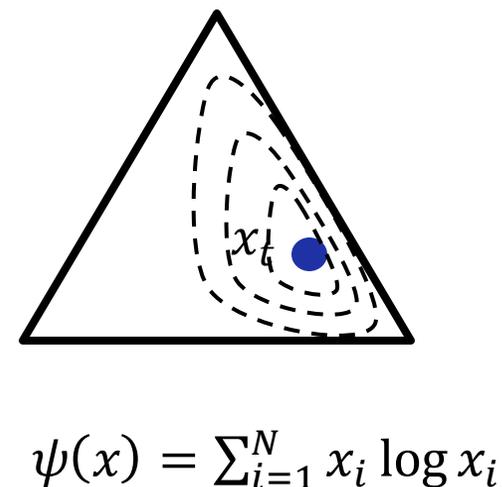
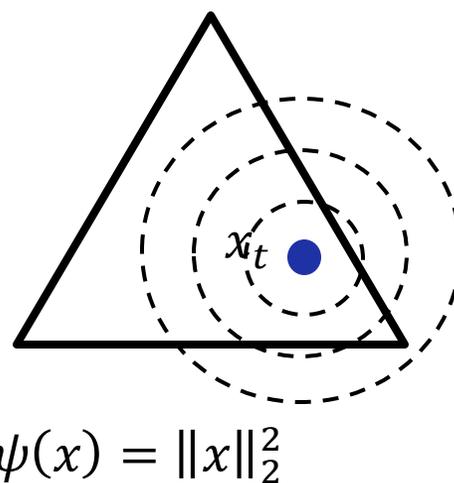
- 簡単のため制約条件がないケースを考えると,

$$g_t + \frac{1}{\eta} (\nabla\psi(x_{t+1}) - \nabla\psi(x_t)) = 0$$

$$\Leftrightarrow \nabla\psi(x_{t+1}) = \nabla\psi(x_t) - \eta g_t \quad \text{写像 } \nabla\psi \text{ で写した世界で勾配降下?}$$



- $X = \Delta(N)$ の場合の $D_\psi(x, x_t)$ の等高線



オンライン鏡像降下法 解析

$$D_\psi(y, x) = \psi(y) - (\psi(x) + \langle \nabla \psi(x), y - x \rangle)$$

$$\text{オンライン鏡像降下法 (OMD)} \quad x_{t+1} = \arg \min_{x \in X} \left\{ \langle g_t, x \rangle + \frac{1}{\eta} D_\psi(x, x_t) \right\}$$

$$\text{定理} \quad \sum_{t=1}^T \langle g_t, x_t - x^* \rangle \leq \frac{1}{\eta} D_\psi(x^*, x_1) + \sum_{t=1}^T \left(\langle g_t, x_t - x_{t+1} \rangle - \frac{1}{\eta} D_\psi(x_{t+1}, x_t) \right) \quad (\forall x^* \in X)$$

(証明) OMD の目的関数の勾配: $g_t + \frac{1}{\eta} (\nabla \psi(x) - \nabla \psi(x_t))$

一次最適性条件より, $\left\langle g_t + \frac{1}{\eta} (\nabla \psi(x_{t+1}) - \nabla \psi(x_t)), x^* - x_{t+1} \right\rangle \geq 0$

$$\Leftrightarrow \langle g_t, x_{t+1} - x^* \rangle \leq \frac{1}{\eta} \langle \nabla \psi(x_{t+1}) - \nabla \psi(x_t), x^* - x_{t+1} \rangle = \frac{1}{\eta} \left(D_\psi(x^*, x_t) - D_\psi(x^*, x_{t+1}) - D_\psi(x_{t+1}, x_t) \right)$$

よって,

$$\begin{aligned} \sum_{t=1}^T \langle g_t, x_t - x^* \rangle &= \sum_{t=1}^T \langle g_t, x_{t+1} - x^* \rangle + \sum_{t=1}^T \langle g_t, x_t - x_{t+1} \rangle \\ &\leq \frac{1}{\eta} \sum_{t=1}^T \left(D_\psi(x^*, x_t) - D_\psi(x^*, x_{t+1}) - D_\psi(x_{t+1}, x_t) \right) + \sum_{t=1}^T \langle g_t, x_t - x_{t+1} \rangle \\ &= \frac{1}{\eta} \left(D_\psi(x^*, x_1) - D_\psi(x^*, x_{T+1}) - \sum_{t=1}^T D_\psi(x_{t+1}, x_t) \right) + \sum_{t=1}^T \langle g_t, x_t - x_{t+1} \rangle \\ &\leq \frac{1}{\eta} D_\psi(x^*, x_1) + \sum_{t=1}^T \left(\langle g_t, x_t - x_{t+1} \rangle - \frac{1}{\eta} D_\psi(x_{t+1}, x_t) \right) \end{aligned}$$

オンライン鏡像降下法 解析

$$D_\psi(y, x) = \psi(y) - (\psi(x) + \langle \nabla \psi(x), y - x \rangle)$$

$$\text{オンライン鏡像降下法 (OMD)} \quad x_{t+1} = \arg \min_{x \in X} \left\{ \langle g_t, x \rangle + \frac{1}{\eta} D_\psi(x, x_t) \right\}$$

$$\text{定理} \quad \sum_{t=1}^T \langle g_t, x_t - x^* \rangle \leq \underbrace{\frac{1}{\eta} D_\psi(x^*, x_1)}_{\text{Penalty 項}} + \underbrace{\sum_{t=1}^T \left(\langle g_t, x_t - x_{t+1} \rangle - \frac{1}{\eta} D_\psi(x_{t+1}, x_t) \right)}_{\text{Stability 項}} \quad (\forall x^* \in X)$$

Penalty 項

Stability 項

- **Penalty 項** : $x_1 \in \arg \min_{x \in X} \psi(x)$ ならば, $D_\psi(x^*, x_1) \leq \psi(x^*) - \psi(x_1) \leq \max_{x \in X} \psi(x) - \min_{x \in X} \psi(x)$
- **Stability 項** : $D_\psi(y, x) \geq c \|y - x\|^2$ ならば,

$$\langle g_t, x_t - x_{t+1} \rangle - \frac{1}{\eta} D_\psi(x_{t+1}, x_t) \leq \|g_t\|_* \cdot \|x_t - x_{t+1}\| - \frac{c}{\eta} \|x_t - x_{t+1}\|^2 \leq \max_{z \geq 0} \left\{ \|g_t\|_* z - \frac{c}{\eta} z^2 \right\} = \frac{\eta}{4c} \|g_t\|_*^2$$

$$(\|y\|_* = \max\{|y^\top x| \mid \|x\| = 1\} = \sup_{x \neq 0} \frac{|y^\top x|}{\|x\|} : \|\cdot\| \text{ の双対ノルム})$$

例

$$\text{OGD} : \psi(x) = \frac{1}{2} \|x\|_2^2 \text{ とすると, } \text{Penalty} \leq \frac{1}{2\eta} \max_{x \in X} \|x\|_2^2 \leq \frac{1}{2\eta} D^2, \text{ Stability} \leq \frac{\eta}{2} \sum_{t=1}^T \|g_t\|_2^2 \leq \frac{\eta}{2} G^2 T$$

$$\text{Hedge} : X = \Delta(N), \psi(x) = \sum_{i=1}^N x_i \log x_i \text{ とすると, } \text{Penalty} \leq \frac{1}{2\eta} \log N,$$

$$D_\psi(y, x) = \text{KL}(y \| x) \geq \frac{1}{2} \|x - y\|_1^2 \quad (\text{ Pinsker の不等式}) \text{ より, } \text{Stability} \leq \frac{\eta}{2} \sum_{t=1}^T \|g_t\|_\infty^2 = O(\eta T)$$

オンライン鏡像降下法（時変ステップサイズ）

$$\text{オンライン鏡像降下法 (OMD)} \quad x_{t+1} = \arg \min_{x \in X} \left\{ \langle g_t, x \rangle + \frac{1}{\eta_t} D_\psi(x, x_t) \right\} \quad (\eta_1 \geq \eta_2 \geq \dots)$$

$$\begin{aligned} \text{定理} \quad \sum_{t=1}^T \langle g_t, x_t - x^* \rangle &\leq \frac{1}{\eta_1} D_\psi(x^*, x_1) + \sum_{t=1}^T \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} \right) D_\psi(x^*, x_{t+1}) \\ &\quad + \sum_{t=1}^T \left(\langle g_t, x_t - x_{t+1} \rangle - \frac{1}{\eta_t} D_\psi(x_{t+1}, x_t) \right) \quad (\forall x^* \in X) \end{aligned}$$

（証明） OMD の目的関数の勾配： $g_t + \frac{1}{\eta_t} (\nabla\psi(x) - \nabla\psi(x_t))$

一次最適性条件より， $\left\langle g_t + \frac{1}{\eta_t} (\nabla\psi(x_{t+1}) - \nabla\psi(x_t)), x^* - x_{t+1} \right\rangle \geq 0$

$$\Leftrightarrow \langle g_t, x_{t+1} - x^* \rangle \leq \frac{1}{\eta_t} \langle \nabla\psi(x_{t+1}) - \nabla\psi(x_t), x^* - x_{t+1} \rangle = \frac{1}{\eta_t} \left(D_\psi(x^*, x_t) - D_\psi(x^*, x_{t+1}) - D_\psi(x_{t+1}, x_t) \right)$$

よって，

$$\begin{aligned} \sum_{t=1}^T \langle g_t, x_t - x^* \rangle &= \sum_{t=1}^T \langle g_t, x_{t+1} - x^* \rangle + \sum_{t=1}^T \langle g_t, x_t - x_{t+1} \rangle \\ &\leq \frac{1}{\eta_t} \sum_{t=1}^T \left(D_\psi(x^*, x_t) - D_\psi(x^*, x_{t+1}) - D_\psi(x_{t+1}, x_t) \right) + \sum_{t=1}^T \langle g_t, x_t - x_{t+1} \rangle \\ &\leq \frac{1}{\eta_1} D_\psi(x^*, x_1) + \sum_{t=1}^T \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} \right) D_\psi(x^*, x_{t+1}) + \sum_{t=1}^T \left(\langle g_t, x_t - x_{t+1} \rangle - \frac{1}{\eta_t} D_\psi(x_{t+1}, x_t) \right) \end{aligned}$$

Follow-the-regularized-leader (FTRL)

$$\text{Follow-the-regularized-leader (FTRL)} \quad x_t = \arg \min_{x \in X} \left\{ \left\langle \sum_{s=1}^{t-1} g_s, x \right\rangle + \frac{1}{\eta} \psi(x) \right\}$$

- 例 1 : $X = \Delta(N)$ (確率単体), $\psi(x) = \sum_{i=1}^N x_i \log x_i$ (Negentropy) のとき Hedge アルゴリズム (\because) Lagrange の未定乗数法より,

$$\sum_{s=1}^{t-1} g_s(i) + \frac{1}{\eta} \log x_t(i) = \text{const} \implies x_t(i) \propto \exp(-\eta \sum_{s=1}^{t-1} g_s(i))$$

- 例 2 : $\psi(x) = \frac{1}{2} \|x\|_2^2$ のとき, "lazy" オンライン勾配降下 :

$$x_t \in \operatorname{argmin}_{x \in X} \left\{ \left\| x + \eta \sum_{s=1}^{t-1} g_s \right\|_2^2 \right\} \iff \tilde{x}_{t+1} = \tilde{x}_t - \eta g_t, \tilde{x}_{t+1} = \operatorname{Proj}_X(\tilde{x}_{t+1})$$

(cf. オンライン勾配降下 : $\tilde{x}_{t+1} = x_t - \eta g_t, \tilde{x}_{t+1} = \operatorname{Proj}_X(\tilde{x}_{t+1})$)

Follow-the-regularized-leader (FTRL)

$$\text{Follow-the-regularized-leader (FTRL)} \quad x_t = \arg \min_{x \in X} \left\{ \left\langle \sum_{s=1}^{t-1} g_s, x \right\rangle + \frac{1}{\eta} \psi(x) \right\}$$

$$\text{定理} \quad \sum_{t=1}^T \langle g_t, x_t - x^* \rangle \leq \frac{1}{\eta} (\psi(x^*) - \psi(x_1)) + \sum_{t=1}^T \left(\langle g_t, x_t - x_{t+1} \rangle - \frac{1}{\eta} D_\psi(x_{t+1}, x_t) \right) \quad (\forall x^* \in X)$$

Follow-the-regularized-leader (FTRL) 時変ステップサイズ

$$x_t = \arg \min_{x \in X} \left\{ \left\langle \sum_{s=1}^{t-1} g_s, x \right\rangle + \frac{1}{\eta_t} \psi(x) \right\}$$

$$\text{定理} \quad \sum_{t=1}^T \langle g_t, x_t - x^* \rangle \leq \frac{1}{\eta_1} (\psi(x^*) - \psi(x_1)) + \sum_{t=1}^T \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} \right) (\psi(x^*) - \psi(x_{t+1})) \\ + \sum_{t=1}^T \left(\langle g_t, x_t - x_{t+1} \rangle - \frac{1}{\eta_t} D_\psi(x_{t+1}, x_t) \right) \quad (\forall x^* \in X)$$

多腕バンディット問題

- $A := [N] := \{1, 2, \dots, N\}$: 有限個の選択肢の集合
- $\ell_t = (\ell_t(i))_{i \in [N]} \in [0, 1]^N$: 損失ベクトル
- $T \in \mathbb{N}$: ラウンド数

- For $t = 1, 2, \dots, T$
 - 環境が損失ベクトル $\ell_t \in [0, 1]^N$ を決定
 - 学習者 (プレイヤー) が $i_t \in [N]$ を決定
 - 学習者は損失 $\ell_t(i_t)$ を被り, 損失値 $\ell_t(i_t)$ を観測
($i \neq i_t$ についての $\ell_t(i)$ は観測できない)

- 弱適応的敵対者モデルを想定: ℓ_t は $\{i_1, i_2, \dots, i_{t-1}\}$ に依存してもよい

$$R_T := \max_{i^* \in [N]} \mathbf{E} \left[\sum_{t=1}^T \ell_t(i_t) - \sum_{t=1}^T \ell_t(i^*) \right]$$

ここで $\mathbf{E}[\cdot]$ は学習者および環境のすべてのランダム性についての期待値

多腕バンディット問題に対するEXP3アルゴリズム

EXP3アルゴリズム

- $\eta > 0$: ステップサイズ, $w_1 = (w_1(i))_{i \in [N]} \in \mathbb{R}^N$ を $w_1(i) = 1 (\forall i \in [N])$ で初期化
- For $t = 1, 2, \dots$
 - $x_t = \frac{w_t}{\|w_t\|_1}$ (x_t は w_t に比例する確率分布)
 - 分布 x_t に従って $i_t \in [N]$ を選択 ($\Pr(i_t = i) = x_t(i)$), $\ell_t(i_t) \in [0, 1]$ を観測
 - ℓ_t の不偏推定量を構成: $\hat{\ell}_t(i_t) = \frac{1}{x_t(i_t)} \ell_t(i_t)$, $\hat{\ell}_t(i) = 0 (\forall i \neq i_t)$
 - $w_{t+1}(i) = w_t(i) \exp(-\eta \hat{\ell}_t(i))$ ($\forall i \in [N]$)

- 大部分は Hedge アルゴリズムと同様
- ℓ_t を観測できないので, 代わりに $\hat{\ell}_t$ を用いる

定理 $\max_{i^* \in [N]} \mathbf{E}[\sum_{t=1}^T \ell_t(i_t) - \sum_{t=1}^T \ell_t(i^*)] \leq \frac{1}{\eta} \log N + \frac{\eta}{2} NT$

したがって $\eta = \Theta\left(\frac{\sqrt{\log N}}{\sqrt{NT}}\right)$ とおくと $R_T = O(\sqrt{TN \log N})$

EXP3アルゴリズムにおける不偏推定量

$\hat{\ell}_t$ の定義：

- 分布 x_t に従って $i_t \in [N]$ を選択 ($\Pr(i_t = i) = x_t(i)$) , $\ell_t(i_t) \in [0,1]$ を観測
- ℓ_t の不偏推定量を構成： $\hat{\ell}_t(i_t) = \frac{1}{x_t(i_t)} \ell_t(i_t)$, $\hat{\ell}_t(i) = 0$ ($\forall i \neq i_t$)

補題 $\hat{\ell}_t$ は ℓ_t の不偏推定量, すなわち $\mathbf{E}_{i_t \sim x_t} [\hat{\ell}_t(i) \mid \ell_t, x_t] = \ell_t(i)$ ($\forall i \in [N]$)

(証明) 指示関数 $\mathbb{I}[\cdot]$ ($\mathbb{I}[\text{true}] = 1$, $\mathbb{I}[\text{false}] = 0$) を用いて $\hat{\ell}_t$ は次のように表記できる：

$$\hat{\ell}_t(i) = \frac{\mathbb{I}[i_t = i]}{x_t(i)} \ell_t(i)$$

よって

$$\mathbf{E}_{i_t \sim x_t} [\hat{\ell}_t(i) \mid \ell_t, x_t] = \mathbf{E}_{i_t \sim x_t} \left[\frac{\mathbb{I}[i_t = i]}{x_t(i)} \ell_t(i) \mid \ell_t, x_t \right] = \frac{\Pr(i_t = i)}{x_t(i)} \ell_t(i) = \ell_t(i)$$

EXP3アルゴリズムにおける不偏推定量

$\hat{\ell}_t$ の定義：

- 分布 x_t に従って $i_t \in [N]$ を選択 ($\Pr(i_t = i) = x_t(i)$) , $\ell_t(i_t) \in [0,1]$ を観測
- ℓ_t の不偏推定量を構成： $\hat{\ell}_t(i_t) = \frac{1}{x_t(i_t)} \ell_t(i_t)$, $\hat{\ell}_t(i) = 0$ ($\forall i \neq i_t$)

補題 $\hat{\ell}_t$ は ℓ_t の不偏推定量, すなわち $\mathbf{E}_{i_t \sim x_t} [\hat{\ell}_t(i) | \ell_t, x_t] = \ell_t(i)$ ($\forall i \in [N]$)

補題より,

$$\mathbf{E} \left[\sum_{t=1}^T \ell_t(i_t) - \sum_{t=1}^T \ell_t(i^*) \right] = \mathbf{E} \left[\sum_{t=1}^T \ell_t^\top x_t - \sum_{t=1}^T \ell_t(i^*) \right] = \mathbf{E} \left[\sum_{t=1}^T \hat{\ell}_t^\top x_t - \sum_{t=1}^T \ell_t(i^*) \right]$$

⇒ エキスパート問題に帰着！

しかし, $\hat{\ell}_t$ が非有界なので注意が必要

$x_t(i_t)$ が 0 に近づくと, $\hat{\ell}_t(i_t)$ はいくらでも大きくなる

EXP3アルゴリズム リグレット上界

- $i_t \sim x_t, \hat{\ell}_t(i_t) = \frac{1}{x_t(i_t)} \ell_t(i_t), \hat{\ell}_t(i) = 0 (\forall i \neq i_t)$
- $x_{t+1} \in \arg \min_{x \in X} \left\{ \langle \hat{\ell}_t, x \rangle + \frac{1}{\eta} D_\psi(x, x_t) \right\}$, where $\psi(x) = \sum_{i=1}^N x_i \log x_i$

定理 $\max_{i^* \in [N]} \mathbf{E}[\sum_{t=1}^T \ell_t(i_t) - \sum_{t=1}^T \ell_t(i^*)] \leq \frac{1}{\eta} \log N + \frac{\eta}{2} NT$

(証明) 鏡像降下法のリグレット上界より,

$$\mathbf{E} \left[\sum_{t=1}^T \hat{\ell}_t^\top x_t - \sum_{t=1}^T \hat{\ell}_t(i^*) \right] \leq \frac{\log N}{\eta} + \mathbf{E} \left[\sum_{t=1}^T \left(\langle \hat{\ell}_t, x_t - x_{t+1} \rangle - \frac{1}{\eta} D_\psi(x_{t+1}, x_t) \right) \right]$$

ここで, $\langle \ell, x - y \rangle - \frac{1}{\eta} D_\psi(y, x) \leq \sum_{i=1}^N x(i) (\exp(-\eta \ell(i)) + \eta \ell(i) - 1)$ より (次ページ),

$$\begin{aligned} \mathbf{E} \left[\langle \hat{\ell}_t, x_t - x_{t+1} \rangle - \frac{1}{\eta} D_\psi(x_{t+1}, x_t) \right] &\leq \frac{1}{\eta} \sum_{i=1}^N \mathbf{E} \left[x_t(i) (\exp(-\eta \hat{\ell}_t(i)) + \eta \hat{\ell}_t(i) - 1) \right] \\ &\leq \frac{\eta}{2} \sum_{i=1}^N \mathbf{E} [x_t(i) \hat{\ell}_t(i)^2] = \frac{\eta}{2} \sum_{i=1}^N \mathbf{E} \left[x_t(i) \cdot \left(\frac{\mathbb{I}[i_t=i]}{x_t(i)} \ell_t(i) \right)^2 \right] \leq \frac{\eta}{2} \sum_{i=1}^N \mathbf{E} \left[\frac{\mathbb{I}[i_t=i]}{x_t(i)} \right] = \frac{\eta N}{2} \end{aligned}$$

Negentropy の stability 項

補題 $\psi(x) = \sum_{i=1}^N x(i) \log x(i)$ のとき,

$$\langle \ell, x - y \rangle - \frac{1}{\eta} D_{\psi}(y, x) \leq \frac{1}{\eta} \sum_{i=1}^N x(i) (\exp(-\eta \ell(i)) + \eta \ell(i) - 1)$$

(証明) $\max_{y \in \mathbb{R}_{>0}^N} \left\{ \langle \ell, x - y \rangle - \frac{1}{\eta} D_{\psi}(y, x) \right\}$ を考えると, この最大値は $\nabla \psi(y) = \nabla \psi(x) - \eta \ell$ すなわち

$y(i) = x(i) \exp(-\eta \ell(i))$ のとき達成される. このとき,

$$\langle \ell, x - y \rangle - \frac{1}{\eta} D_{\psi}(y, x) = \frac{1}{\eta} \sum_{i=1}^N x(i) \left(\frac{y(i)}{x(i)} + \log \frac{x(i)}{y(i)} - 1 \right) = \frac{1}{\eta} \sum_{i=1}^N x(i) (\exp(-\eta \ell(i)) + \eta \ell(i) - 1)$$

Negentropy の stability 項

補題 $\psi(x) = \sum_{i=1}^N x(i) \log x(i)$ のとき,

$$\langle \ell, x - y \rangle - \frac{1}{\eta} D_{\psi}(y, x) \leq \frac{1}{\eta} \sum_{i=1}^N x(i) (\exp(-\eta \ell(i)) + \eta \ell(i) - 1)$$

- $z \leq 1 \Rightarrow \exp(z) \leq 1 + z + z^2$ を使うと,

$$\forall i, \eta \ell(i) \geq -1 \Rightarrow \langle \ell, x - y \rangle - \frac{1}{\eta} D_{\psi}(y, x) = O\left(\frac{1}{\eta} \sum_{i=1}^N x(i) (\eta \ell(i))^2\right) = O(\eta \sum_{i=1}^N x(i) \ell(i)^2)$$

このバウンドは ℓ が非負なら常に有効だが、 マイナス方向に大きすぎると使えないことに注意

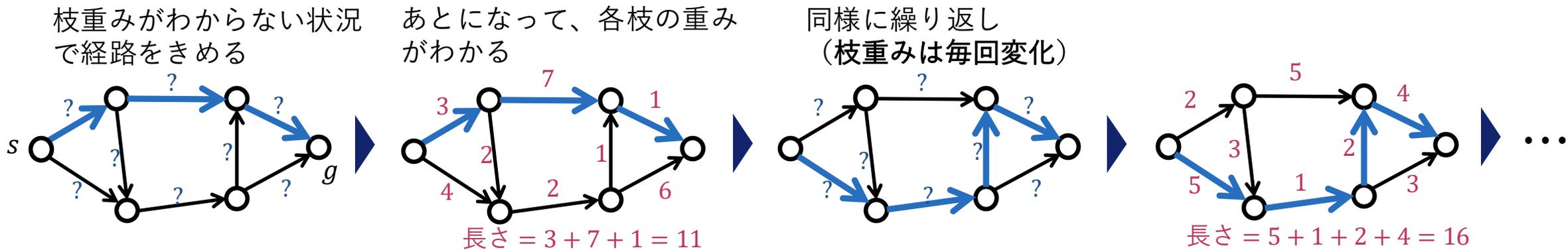
講義の概要

- リグレット解析の枠組みとその応用
 - オンライン凸最適化問題とオンライン勾配降下法
 - Online-to-batch 変換
 - ミニマックス定理
 - エクスパート問題と Hedge アルゴリズム
- 正則化を用いたアルゴリズムとリグレット解析
 - オンライン鏡像降下法
 - 多腕バンディット問題とEXP3アルゴリズム
- オンライン組合せ最適化
 - 完全情報フィードバック
 - セミバンディットフィードバック
 - バンディットフィードバック

完全情報 オンライン組合せ最適化

オンライン最短経路問題: [AK04], [AK08]

- 事前に重み無し有向グラフ $G = (V, E)$ と始点 $s \in V$, 終点 $g \in V$ が与えられる
- 各ラウンド $t = 1, 2, \dots, T$ において,
経路 P_t を決定したあとで, 枝重み (通行所要時間など) $w_t: E \rightarrow \mathbb{R}_{\geq 0}$ を観測

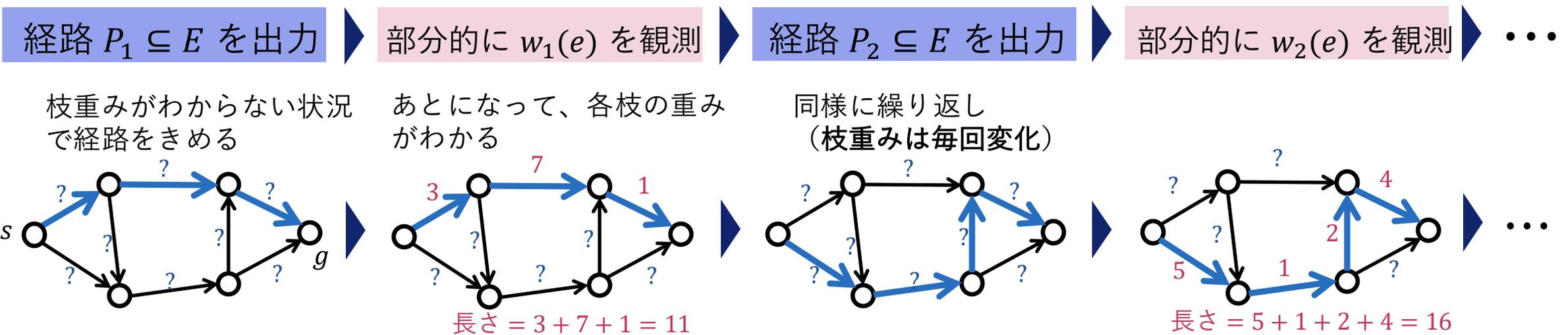


完全情報: 選ばなかった枝の重みも全て観測可能

セミバンディット オンライン組合せ最適化

オンライン最短経路問題: [AK04], [AK08]

- 事前に重み無し有向グラフ $G = (V, E)$ と始点 $s \in V$, 終点 $g \in V$ が与えられる
- 各ラウンド $t = 1, 2, \dots, T$ において,
経路 P_t を決定したあとで, P_t に含まれる e について $w_t(e)$ を観測

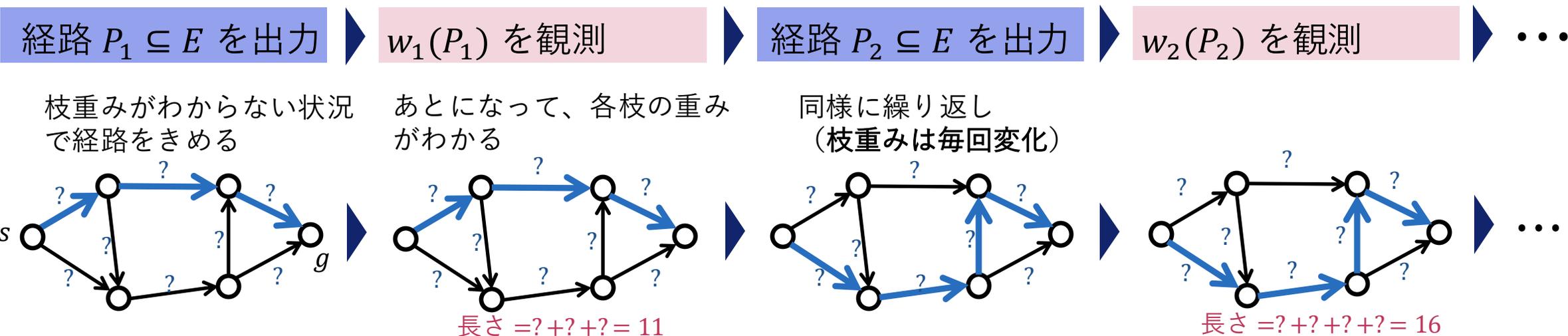


セミバンディット：選んだ枝の重みのみ観測可能

(フル) バンディット オンライン組合せ最適化

オンライン最短経路問題: [AK04], [AK08]

- 事前に重み無し有向グラフ $G = (V, E)$ と始点 $s \in V$, 終点 $g \in V$ が与えられる
- 各ラウンド $t = 1, 2, \dots, T$ において,
経路 P_t を決定したあとで, w_t についての P_t の経路長 $w_t(P_t)$ を観測



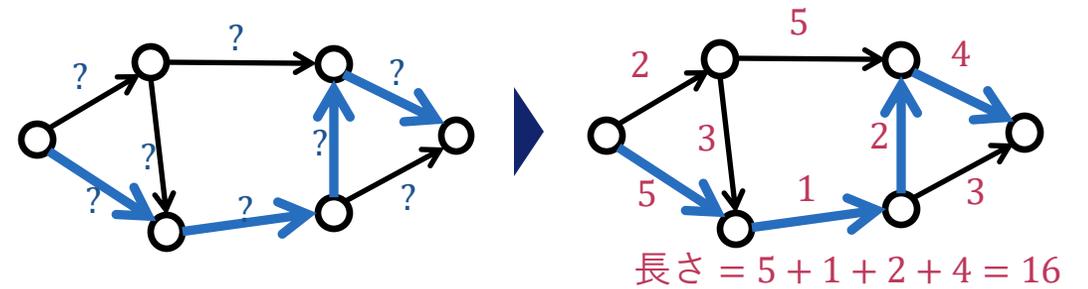
セミバンディット：選んだ経路の経路長=重みの総和のみ観測可能

問題設定の分類

	$A = [N]$	$A \subseteq \{0,1\}^N \cong 2^{[N]}$
完全情報	エキスパート問題 <ul style="list-style-type: none"> • OGD: $R_T = O(\sqrt{TN})$ • Hedge: $R_T = O(\sqrt{T \log N})$ 	完全情報オンライン組合せ最適化 <ul style="list-style-type: none"> • Hedge: • OMD: • FTPL:
部分観測	多腕バンディット <ul style="list-style-type: none"> • EXP3: $R_T = O(\sqrt{TN \log N})$ 	セミバンディット組合せ最適化 <ul style="list-style-type: none"> • OMD:
		バンディット組合せ最適化 <ul style="list-style-type: none"> • EXP2:

(完全情報) オンライン組合せ最適化問題

- $A \subseteq \{0,1\}^N$: 実行可能領域 $[N]$ の部分集合族に対応
- $m := \max_{a \in A} \|a\|_1$
- $\ell_t = (\ell_t(i))_{i \in [N]} \in [0,1]^N$: 損失ベクトル
- $T \in \mathbb{N}$: ラウンド数



- For $t = 1, 2, \dots, T$
 - 環境が損失ベクトル $\ell_t \in [0,1]^N$ を決定
 - 学習者 (プレイヤー) が $a_t \in A$ を決定
 - 学習者は損失 $\ell_t^\top a_t$ を被り, 損失ベクトル ℓ_t を観測
- 弱適応的敵対者モデルを想定: ℓ_t は $\{a_1, a_2, \dots, a_{t-1}\}$ に依存してもよい
- リグレット R_T : 学習者の意思決定の「良さ」の指標

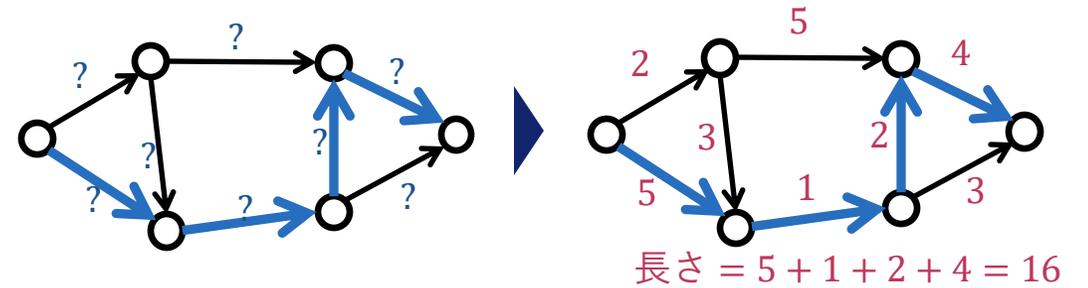
$$R_T := \max_{a^* \in A} \mathbf{E} \left[\sum_{t=1}^T \ell_t^\top a_t - \sum_{t=1}^T \ell_t^\top a^* \right]$$

(完全情報) オンライン組合せ最適化問題

- $A \subseteq \{0,1\}^N$: 実行可能領域 $[N]$ の部分集合族に対応
- $m := \max_{a \in A} \|a\|_1$
- $\ell_t = (\ell_t(i))_{i \in [N]} \in [0,1]^N$: 損失ベクトル

• For $t = 1, 2, \dots, T$

- 環境が損失ベクトル $\ell_t \in [0,1]^N$ を決定
- 学習者 (プレイヤー) が $a_t \in A$ を決定
- 学習者は損失 $\ell_t^\top a_t$ を被り, 損失ベクトル ℓ_t を観測



• 典型的な仮定:

$\ell \in \mathbb{R}^N$ が与えられたとき $O(\ell) \in \arg \min_{a \in A} \{\ell^\top a\}$ を返す線形最適化オラクルが

与えられていることを仮定する

Online-to-batch 変換の結果を考えると, リーズナブルな仮定

(完全情報) オンライン組合せ最適化

- $A \subseteq \{0,1\}^N$: 実行可能領域
- $m := \max_{a \in A} \|a\|_1$

アプローチ 1 : A 上の Hedge アルゴリズム

アプローチ 2 : A の凸包上の オンライン凸最適化

アプローチ 3 : Follow-the-perturbed-leader (FTPL)

(完全情報) オンライン組合せ最適化

アプローチ 1 : A 上の Hedge アルゴリズム

- $A \subseteq \{0,1\}^N$:
- $m := \max_{a \in A} \|a\|_1$

Hedge アルゴリズム / 乗算型重み更新 (MWU)

- 入力 : $\eta > 0$: ステップサイズ
- $w_1 = (w_1(a))_{a \in A} \in \mathbb{R}^A$ を $w_1(a) = 1 (\forall a \in A)$ で初期化
- For $t = 1, 2, \dots$
 - $x_t = \frac{w_t}{\|w_t\|_1}$ (x_t は w_t に比例する A 上の確率分布)
 - a_t : 分布 x_t からサンプリング
 - $w_{t+1}(a) = w_t(a) \exp(-\eta \ell_t^\top a)$ ($\forall i \in [N]$)

定理 $0 \leq \ell_t^\top a \leq c$ ($\forall t, \forall a \in A$) ならば, $\max_{a^* \in [N]} \mathbf{E}[\sum_{t=1}^T \ell_t^\top a_t - \sum_{t=1}^T \ell_t^\top a^*] \leq \frac{1}{\eta} \log |A| + \frac{\eta}{2} c^2 T$

したがって $\eta = \Theta\left(\frac{\sqrt{\log |A|}}{c\sqrt{T}}\right)$ とおくと $R_T = O(c\sqrt{T \log |A|}) = O(c\sqrt{Tm \log N})$

リグレット上界は良好だが, 分布 x_t ($x_t(a) \propto \exp(-\eta \sum_{s=1}^{t-1} \ell_s^\top a)$) からのサンプリングが必要
ナイーブな実装では $O(m|A|)$ 時間必要

(完全情報) オンライン組合せ最適化

- $A \subseteq \{0,1\}^N$: 実行可能領域
- $m := \max_{a \in A} \|a\|_1$

アプローチ 1 : A 上の Hedge アルゴリズム

- $R_T = O(m\sqrt{T \log |A|}) = O(\sqrt{m^3 T \log N})$
- A 上のサンプリングが必要

アプローチ 2 : A の凸包上の オンライン凸最適化

アプローチ 3 : Follow-the-perturbed-leader (FTPL)

(完全情報) オンライン組合せ最適化

アプローチ 2 : A の凸包上のオンライン凸最適化

- $A \subseteq \{0,1\}^N$:
- $m := \max_{a \in A} \|a\|_1$

- $X \subseteq [0,1]^N$: A の凸包とし, X 上のオンライン凸最適化に帰着

凸包上のオンライン勾配降下

- $\eta > 0$: ステップサイズ, $x_1 \in X$: 初期解
- For $t = 1, 2, \dots$
 - $\mathbf{E}[a_t] = x_t$ が成り立つように $a_t \in A$ を選択
 - $\tilde{x}_{t+1} = x_t - \eta \ell_t$
 - $x_{t+1} = \text{Proj}_X(\tilde{x}_{t+1}) \in \arg \min_{x \in X} \|x - \tilde{x}_{t+1}\|_2$

定理 $\|\ell_t\|_2 \leq G$ ($\forall t$) ならば, $R_T := \max_{a^* \in [N]} \mathbf{E}[\sum_{t=1}^T \ell_t^\top a_t - \sum_{t=1}^T \ell_t^\top a^*] \leq \frac{1}{2} \eta G^2 T + \frac{m}{\eta}$

したがって $\eta = \Theta\left(\frac{\sqrt{m}}{G\sqrt{T}}\right)$ とおくと $R_T = O(G\sqrt{mT})$

(完全情報) オンライン組合せ最適化

アプローチ 2 : A の凸包上のオンライン凸最適化

- $A \subseteq \{0,1\}^N$
- $X = \text{Conv}(A)$

- $\mathbf{E}[a_t] = x_t$ が成り立つように $a_t \in A$ を選択
- $\sum_{j=1}^J \lambda_j a_j \approx x_t, \sum_{j=1}^J \lambda_j = 1$ を満たす $\{\lambda_j\} \subseteq [0,1], \{a_j\} \subseteq A$ を求め、確率 λ_j で a_j を出力すればよい
- A 上の線形最適化オラクルがあれば、多項式時間で $\{\lambda_j\}, \{a_j\}$ を計算可能だが、複雑 (cf. Corollary 14.1g in [A. Schrijver. Theory of Linear and Integer Programming. 1998])
- ミニマックス最適化に基づく方法：所与の $x' = x_t \in X$ に対して、 $Y = \{y \in \mathbb{R}^N \mid \|y\| \leq 1\}, F: X \times Y \rightarrow \mathbb{R}, F(x, y) = (x - x')^\top y$ とすると、
 - $\max_{y \in Y} F(x, y) = \|x - x'\|, \min_{x \in X} \max_{y \in Y} F(x, y) = 0$
 - $a_s = \mathcal{O}(y_s) \in \arg \min_{a \in A} F(a, y_s) = \arg \min_{a \in A} \{y_s^\top a\}$
 - $y_s : f_s(y) := -F(a_s, y)$ に対するオンライン凸最適化
 - $\bar{a}_{T'} = \frac{1}{T'} \sum_{s=1}^{T'} a_s$ とし、 $\{y_s\}$ のリグレットを $R_{T'}$ とすると

$$\max_{y \in Y} F(\bar{a}_{T'}, y) - \min_{x \in X} \max_{y \in Y} F(x, y) = \|\bar{a}_{T'} - x'\| \leq \frac{R_{T'}}{T'} \rightarrow_{T' \rightarrow \infty} 0$$

(完全情報) オンライン組合せ最適化

- $A \subseteq \{0,1\}^N$: 実行可能領域
- $m := \max_{a \in A} \|a\|_1$

アプローチ 1 : A 上の Hedge アルゴリズム

- $R_T = O(m\sqrt{T \log |A|}) = O(\sqrt{m^3 T \log N})$
- A 上のサンプリングが必要

アプローチ 2 : A の凸包上の オンライン凸最適化

- $R_T = O(\sqrt{nmT})$
- A 上の線形最適化オラクルを用いて実行可能

アプローチ 3 : Follow-the-perturbed-leader (FTPL)

(完全情報) オンライン組合せ最適化

アプローチ 3 : Follow-the-perturbed-leader

- $A \subseteq \{0,1\}^N$
- $X = \text{Conv}(A)$

Follow-the-regularized-leader (FTRL)

- \mathbb{R}^N 上の適当な分布 \mathcal{D} を設定
- For $t = 1, 2, \dots$
 - r_t を \mathcal{D} からサンプリングし,
 - $a_t = \mathcal{O}(\sum_{s=1}^{t-1} \ell_s + r_t) \in \arg \min_{a \in A} \left\{ \langle \sum_{s=1}^{t-1} \ell_s, a \rangle + \frac{1}{\eta} \langle r_t, a \rangle \right\}$

定理 $\mathcal{D} : [0,1]^N$ 上の一様分布とすると,

$$\mathbf{E} \left[\sum_{t=1}^T \ell_t^\top a_t - \sum_{t=1}^T \ell_t^\top a^* \right] \leq \frac{m}{\eta} + \eta m \sum_{t=1}^T \|\ell_t\|_\infty \|\ell_t\|_1$$

(完全情報) オンライン組合せ最適化

アプローチ 3 : Follow-the-perturbed-leader

- $A \subseteq \{0,1\}^N$
- $X = \text{Conv}(A)$

Follow-the-regularized-leader (FTRL)

- $r_t \sim \mathcal{D}, a_t = \mathcal{O} \left(\sum_{s=1}^{t-1} \ell_s + \frac{1}{\eta} r_t \right) \in \arg \min_{a \in A} \left\{ \left\langle \sum_{s=1}^{t-1} \ell_s, a \right\rangle + \frac{1}{\eta} \langle r_t, a \rangle \right\}$

定理 $\mathcal{D} : [0,1]^N$ 上の一様分布とすると,

$$\mathbf{E}[\sum_{t=1}^T \ell_t^\top a_t - \sum_{t=1}^T \ell_t^\top a^*] \leq \frac{m}{\eta} + \eta m \sum_{t=1}^T \|\ell_t\|_\infty \|\ell_t\|_1$$

(証明) $r \sim \mathcal{D}$: ほかの確率変数とは独立に \mathcal{D} に従う確率変数とする

$$z_t(r) = \mathcal{O} \left(\sum_{s=1}^{t-1} \ell_s + \frac{1}{\eta} r \right), x_t = \mathbf{E}_{r \sim \mathcal{D}}[z_t(r)] \text{ とすると, } \mathbf{E}[\sum_{t=1}^T \ell_t^\top a_t - \sum_{t=1}^T \ell_t^\top a^*] = \mathbf{E}[\sum_{t=1}^T \ell_t^\top x_t - \sum_{t=1}^T \ell_t^\top a^*]$$

$$\left\langle \sum_{t=1}^T \ell_t + \frac{1}{\eta} r, a^* \right\rangle \geq \left\langle \sum_{t=1}^T \ell_t + \frac{1}{\eta} r, z_{T+1}(r) \right\rangle = \left\langle \sum_{t=1}^{T-1} \ell_t + \frac{1}{\eta} r, z_{T+1}(r) \right\rangle + \ell_T^\top z_{T+1}(r)$$

$$\geq \left\langle \sum_{t=1}^{T-1} \ell_t + \frac{1}{\eta} r, z_T(r) \right\rangle + \ell_T^\top z_{T+1}(r) \geq \dots \geq \frac{1}{\eta} \langle r, z_1(r) \rangle + \sum_{t=1}^T \ell_t^\top z_{t+1}(r) \quad r \sim \mathcal{D} \text{ について期待値をとると,}$$

$$\mathbf{E}[\left\langle \sum_{t=1}^T \ell_t, a^* \right\rangle] \geq \mathbf{E} \left[\sum_{t=1}^T \ell_t^\top x_{t+1} + \frac{1}{\eta} \langle r, z_1(r) - a^* \rangle \right] \geq \mathbf{E}[\sum_{t=1}^T \ell_t^\top x_{t+1}] - \frac{m}{\eta}$$

$$\text{よって } \mathbf{E}[\sum_{t=1}^T \ell_t^\top a_t - \sum_{t=1}^T \ell_t^\top a^*] \leq \mathbf{E}[\sum_{t=1}^T (\ell_t^\top x_t - \ell_t^\top x_{t+1})] + \frac{m}{\eta} \quad (\text{つぎのページにつづく})$$

(完全情報) オンライン組合せ最適化

アプローチ 3 : Follow-the-perturbed-leader

- $A \subseteq \{0,1\}^N$
- $X = \text{Conv}(A)$

Follow-the-regularized-leader (FTRL)

- $r_t \sim \mathcal{D}, a_t = \mathcal{O}\left(\sum_{s=1}^{t-1} \ell_s + \frac{1}{\eta} r_t\right) \in \arg \min_{a \in A} \left\{ \langle \sum_{s=1}^{t-1} \ell_s, a \rangle + \frac{1}{\eta} \langle r_t, a \rangle \right\}$

定理 $\mathcal{D} : [0,1]^N$ 上の一様分布とすると,

$$\mathbf{E}[\sum_{t=1}^T \ell_t^\top a_t - \sum_{t=1}^T \ell_t^\top a^*] \leq \frac{m}{\eta} + \eta m \sum_{t=1}^T \|\ell_t\|_\infty \|\ell_t\|_1$$

(証明) $r \sim \mathcal{D}$: ほかの確率変数とは独立に \mathcal{D} に従う確率変数とする

$z_t(r) = \mathcal{O}\left(\sum_{s=1}^{t-1} \ell_s + \frac{1}{\eta} r\right)$, $x_t = \mathbf{E}_{r \sim \mathcal{D}}[z_t(r)]$ とすると, $\mathbf{E}[\sum_{t=1}^T \ell_t^\top a_t - \sum_{t=1}^T \ell_t^\top a^*] \leq \mathbf{E}[\sum_{t=1}^T (\ell_t^\top x_t - \ell_t^\top x_{t+1})] + \frac{m}{\eta}$

$e_t: \mathbb{R}^N \mapsto [0, m\|\ell_t\|_\infty]$ を $e_t(z) = \ell_t^\top \mathcal{O}(z)$ で定義する. 確率変数 Z_t を $Z_t = \sum_{s=1}^{t-1} \ell_s + \frac{1}{\eta} r$ で定めたとき,

Z_t と Z_{t+1} の間の総変動距離は $\text{TV}(Z_t, Z_{t+1}) \leq \eta \|\ell_t\|_1$ を満たすから,

$$\ell_t^\top x_t - \ell_t^\top x_{t+1} = \mathbf{E}[e_t(Z_t)] - \mathbf{E}[e_t(Z_{t+1})] \leq m\|\ell_t\|_\infty \cdot \text{TV}(Z_t, Z_{t+1}) \leq \eta m \|\ell_t\|_\infty \|\ell_t\|_1$$

(完全情報) オンライン組合せ最適化

- $A \subseteq \{0,1\}^N$: 実行可能領域
- $m := \max_{a \in A} \|a\|_1$

アプローチ 1 : A 上の Hedge アルゴリズム

- $R_T = O(m\sqrt{T \log |A|}) = O(\sqrt{m^3 T \log N})$
- A 上のサンプリングが必要

アプローチ 2 : A の凸包上の オンライン凸最適化

- $R_T = O(\sqrt{nmT})$
- A 上の線形最適化オラクルを用いて実行可能

アプローチ 3 : Follow-the-perturbed-leader (FTPL)

- $R_T = O(\sqrt{nm^2T})$
- A 上の線形最適化オラクルを用いて実行可能, 高速かつシンプル

問題設定の分類

$$m := \max_{a \in A} \|a\|_1$$

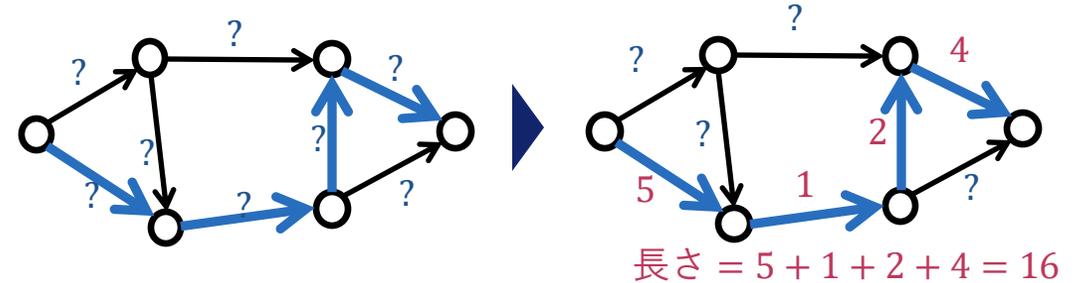
	$A = [N]$	$A \subseteq \{0,1\}^N \cong 2^{[N]}$
完全情報	エキスパート問題 <ul style="list-style-type: none"> • OGD: $R_T = O(\sqrt{TN})$ • Hedge: $R_T = O(\sqrt{T \log N})$ 	完全情報オンライン組合せ最適化 <ul style="list-style-type: none"> • Hedge: $O(\sqrt{m^3 T \log N})$ • OMD: $O(\sqrt{nmT})$ • FTPL: $O(\sqrt{nm^2 T})$
部分観測	多腕バンディット <ul style="list-style-type: none"> • EXP3: $R_T = O(\sqrt{TN \log N})$ 	セミバンディット組合せ最適化 <ul style="list-style-type: none"> • OMD:
		バンディット組合せ最適化 <ul style="list-style-type: none"> • EXP2:

セミバンディット オンライン組合せ最適化

- $A \subseteq \{0,1\}^N$: 実行可能領域 $[N]$ の部分集合族に対応
- $m := \max_{a \in A} \|a\|_1$
- $\ell_t = (\ell_t(i))_{i \in [N]} \in [0,1]^N$: 損失ベクトル

For $t = 1, 2, \dots, T$

- 環境が損失ベクトル $\ell_t \in \mathbb{R}_{\geq 0}^N$ を決定
- 学習者（プレイヤー）が $a_t \in A$ を決定
- 学習者は損失 $\ell_t^\top a_t$ を被り, $a_t(i) = 1$ なる i について損失値 $\ell_t(i)$ を観測



- 弱適応的敵対者モデルを想定： ℓ_t は $\{a_1, a_2, \dots, a_{t-1}\}$ に依存してもよい
- リグレット R_T : 学習者の意思決定の「良さ」の指標

$$R_T := \max_{a^* \in A} \mathbf{E} \left[\sum_{t=1}^T \ell_t^\top a_t - \sum_{t=1}^T \ell_t^\top a^* \right]$$

セミバンディット オンライン組合せ最適化 鏡像降下法アプローチ

- $A \subseteq \{0,1\}^N$:
- $m := \max_{a \in A} \|a\|_1$

- $X \subseteq [0,1]^N$: A の凸包とし, X 上のオンライン凸最適化に帰着

凸包上のオンライン鏡像降下

- $\eta > 0$: ステップサイズ, $x_1 \in \arg \min_{x \in X} \psi(x)$: 初期解
- For $t = 1, 2, \dots$
 - $\mathbf{E}[a_t] = x_t$ が成り立つように $a_t \in A$ を選択
 - 不偏推定量 $\hat{\ell}_t \in \mathbb{R}_{\geq 0}^N$ を $\hat{\ell}_t(i) = \frac{\mathbb{I}[a_t(i)=1]}{x_t(i)} \ell_t(i)$ で定義
 - $x_{t+1} \in \arg \min_{x \in X} \left\{ \langle \hat{\ell}_t, x \rangle + \frac{1}{\eta} D_\psi(x, x_t) \right\}$ ($\psi(x) = \sum_{i=1}^N x_i \log x_i$)

定理 $\ell_t \in [0,1]^N$ ($\forall t$) ならば, $R_T := \max_{a^* \in [N]} \mathbf{E}[\sum_{t=1}^T \ell_t^\top a_t - \sum_{t=1}^T \ell_t^\top a^*] \leq \frac{m \log N}{\eta} + \frac{\eta}{2} NT$

したがって $\eta = \Theta\left(\frac{\sqrt{m \log N}}{\sqrt{NT}}\right)$ とおくと $R_T = O(\sqrt{mTN \log N})$

セミバンディット オンライン組合せ最適化 鏡像降下法アプローチ

- $A \subseteq \{0,1\}^N$
- $m := \max_{a \in A} \|a\|_1$
- $X = \text{Conv}(A)$

- $\mathbf{E}[a_t] = x_t$ が成り立つように $a_t \in A$ を選択, $\hat{\ell}_t(i) = \frac{\mathbb{I}[a_t(i)=1]}{x_t(i)} \ell_t(i)$ で定義
- $x_1 \in \arg \min_{x \in X} \psi(x)$, $x_{t+1} \in \arg \min_{x \in X} \left\{ \langle \hat{\ell}_t, x \rangle + \frac{1}{\eta} D_\psi(x, x_t) \right\}$ ($\psi(x) = \sum_{i=1}^N x(i) \log x(i)$)

定理 $\ell_t \in [0,1]^N$ のもとで, $\mathbf{E}[\sum_{t=1}^T \ell_t^\top a_t - \sum_{t=1}^T \ell_t^\top a^*] \leq \frac{m \log N}{\eta} + \frac{\eta}{2} NT$

(証明) $\hat{\ell}_t$ は ℓ_t の不偏推定量 ($\mathbf{E}_{a_t}[\hat{\ell}_t(i) \mid \ell_t, x_t] = \ell_t(i)$) だから,

$$\mathbf{E}[\sum_{t=1}^T \ell_t^\top a_t - \sum_{t=1}^T \ell_t^\top a^*] = \mathbf{E}[\sum_{t=1}^T \hat{\ell}_t^\top x_t - \sum_{t=1}^T \hat{\ell}_t^\top a^*]$$

鏡像降下法のリグレット上界より,

$$\mathbf{E}[\sum_{t=1}^T \hat{\ell}_t^\top x_t - \sum_{t=1}^T \hat{\ell}_t^\top a^*] \leq \frac{\psi(a^*) - \psi(x_1)}{\eta} + \mathbf{E} \left[\sum_{t=1}^T \left(\langle \hat{\ell}_t, x_t - x_{t+1} \rangle - \frac{1}{\eta} D_\psi(x_{t+1}, x_t) \right) \right]$$

$$\psi(a^*) - \psi(x_1) = -\sum_{i=1}^N x_1(i) \log x_1(i) \leq \|x_1\|_1 \log \frac{N}{\|x_1\|_1} \leq m \log N$$

$$\mathbf{E} \left[\langle \hat{\ell}_t, x_t - x_{t+1} \rangle - \frac{1}{\eta} D_\psi(x_{t+1}, x_t) \right] \leq \frac{1}{2} \sum_{i=1}^N \mathbf{E} \left[\frac{\mathbb{I}[a_t(i)=1]}{x_t(i)} \right] \leq \frac{N}{2} \quad (\text{EXP3 のときと同様})$$

問題設定の分類

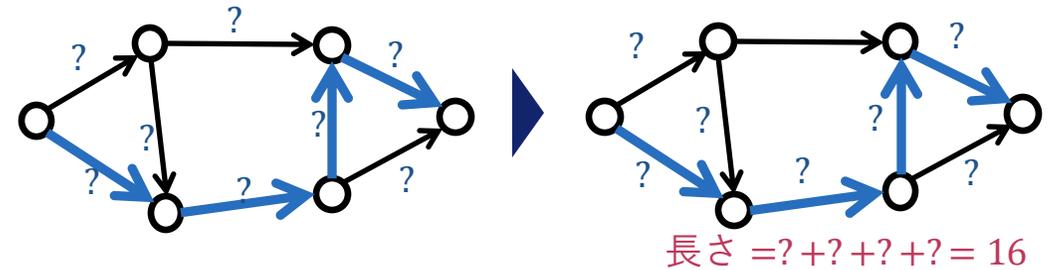
$$m := \max_{a \in A} \|a\|_1$$

	$A = [N]$	$A \subseteq \{0,1\}^N \cong 2^{[N]}$
完全情報	エキスパート問題 <ul style="list-style-type: none"> • OGD: $R_T = O(\sqrt{TN})$ • Hedge: $R_T = O(\sqrt{T \log N})$ 	完全情報オンライン組合せ最適化 <ul style="list-style-type: none"> • Hedge: $O(\sqrt{m^3 T \log N})$ • OMD: $O(\sqrt{nmT})$ • FTPL: $O(\sqrt{nm^2 T})$
部分観測	多腕バンディット <ul style="list-style-type: none"> • EXP3: $R_T = O(\sqrt{TN \log N})$ 	セミバンディット組合せ最適化 <ul style="list-style-type: none"> • OMD: $O(\sqrt{mTN \log N})$
		バンディット組合せ最適化 <ul style="list-style-type: none"> • EXP2:

バンディット オンライン組合せ最適化

- $A \subseteq \{0,1\}^N$: 実行可能領域 $[N]$ の部分集合族に対応
- $m := \max_{a \in A} \|a\|_1$
- $\ell_t = (\ell_t(i))_{i \in [N]} \in [0,1]^N$: 損失ベクトル

- For $t = 1, 2, \dots, T$
 - 環境が損失ベクトル $\ell_t \in \mathbb{R}_{\geq 0}^N$ を決定
 - 学習者 (プレイヤー) が $a_t \in A$ を決定
 - 学習者は損失 $\ell_t^\top a_t$ を被り, $\ell_t^\top a_t$ を観測



- 弱適応的敵対者モデルを想定: ℓ_t は $\{a_1, a_2, \dots, a_{t-1}\}$ に依存してもよい
- リグレット R_T : 学習者の意思決定の「良さ」の指標

$$R_T := \max_{a^* \in A} \mathbf{E} \left[\sum_{t=1}^T \ell_t^\top a_t - \sum_{t=1}^T \ell_t^\top a^* \right]$$

バンディット オンライン組合せ最適化 アプローチと課題

不偏推定量の構成：

- A 上の確率分布 $x_t \in \Delta(A)$ に従って a_t を選択し ($\Pr(a_t = a) = x_t(a)$) $a_t^\top \ell_t$ を観測
- $\hat{\ell}_t \in \mathbb{R}^N$ を次の式で定義：

$$\hat{\ell}_t = S_t^{-1} a_t a_t^\top \ell_t, \quad \text{where } S_t := \mathbf{E}_{a \sim x_t} [a a^\top] = \sum_{a \in A} x_t(a) a a^\top$$

- これは ℓ_t の不偏推定量. 実際, $\mathbf{E}_{a_t \sim x_t} [\hat{\ell}_t \mid x_t, \ell_t] = S_t^{-1} \mathbf{E}_{a \sim x_t} [a a^\top] \ell_t = S_t^{-1} S_t \ell_t = \ell_t$

不偏推定量の課題点：

- $\hat{\ell}_t^\top a$ は非負とは限らない. 分布 x_t によっては, マイナス方向に非有界
- これは Hedge や鏡像降下を用いるときの stability 項を抑えるときに不都合

解決策： G-optimal design を用いる

G-optimal design

$A \subseteq \{0,1\}^N$ は \mathbb{R}^N を張ると仮定する

A 上の分布 $\pi \in \Delta(A)$ に対して次の行列, 関数を定義

- $S(\pi) := \mathbf{E}_{a \sim \pi} [aa^\top] = \sum_{a \in A} \pi(a)aa^\top \in \mathbb{R}^{N \times N}$
- $G(\pi) = \max_{a \in A} a^\top S(\pi)^{-1} a \in \mathbb{R}$

定理 $\min_{\pi \in \Delta(A)} G(\pi) = N$

(略証) ミニマックス定理を用いて示せる (※下記は細部を省略しているので注意)

$X = \Delta(A)$, $Y = \text{Conv}(\{aa^\top \in \mathbb{R}^{N \times N} \mid a \in A\})$,

$F: X \times Y \rightarrow \mathbb{R}$, $F(\pi, U) = \text{tr}(S(\pi)^{-1}U)$ とすると, F は凸-凹関数 (※非自明)

Sion のミニマックス定理より,

$$\min_{\pi \in \Delta(A)} G(\pi) = \min_{\pi \in \Delta(A)} \max_{U \in Y} F(\pi, U) = \max_{U \in Y} \min_{\pi \in \Delta(A)} F(\pi, U) = N$$

$$\because \max_{U \in Y} F(\pi, U) \geq \text{tr}(S(\pi)^{-1}S(\pi)) = N, \quad \min_{\pi \in \Delta(A)} F(\pi, U) \leq \text{tr}(U^{-1}U) = N$$

G-optimal design

$A \subseteq \{0,1\}^N$ は \mathbb{R}^N を張ると仮定する

A 上の分布 $\pi \in \Delta(A)$ に対して次の行列, 関数を定義

- $S(\pi) := \mathbf{E}_{a \sim \pi} [aa^\top] = \sum_{a \in A} \pi(a)aa^\top \in \mathbb{R}^{N \times N}$
- $G(\pi) = \max_{a \in A} a^\top S(\pi)^{-1} a \in \mathbb{R}$

定理 $\min_{\pi \in \Delta(A)} G(\pi) = N$

- この最小値を達成する分布 π^* を A に対する G-optimal design という
- 補足 1 : もともとは実験計画の文脈で導入
- 補足 2 : $\operatorname{argmin}_{\pi \in \Delta(A)} \{G(\pi)\} = \operatorname{argmax}_{\pi \in \Delta(A)} \{\det G(\pi)\}$

これを証明すると同時に定理を示すこともできる

バンディット オンライン組合せ最適化 アプローチと課題

不偏推定量の構成：

- A 上の確率分布 $x_t \in \Delta(A)$ に従って a_t を選択し ($\Pr(a_t = a) = x_t(a)$) $a_t^\top \ell_t$ を観測
- $\hat{\ell}_t \in \mathbb{R}^N$ を次の式で定義：

$$\hat{\ell}_t = S_t^{-1} a_t a_t^\top \ell_t, \quad \text{where } S_t := \mathbf{E}_{a \sim x_t} [a a^\top] = \sum_{a \in A} x_t(a) a a^\top$$

- これは ℓ_t の不偏推定量. 実際, $\mathbf{E}_{a_t \sim x_t} [\hat{\ell}_t | x_t, \ell_t] = S_t^{-1} \mathbf{E}_{a_t \sim x_t} [a_t a_t^\top] \ell_t = S_t^{-1} S_t \ell_t = \ell_t$

不偏推定量の課題点：

- $\hat{\ell}_t^\top a$ は非負とは限らない. 分布 x_t によっては, マイナス方向に非有界
- これは Hedge や鏡像降下を用いるときの stability 項を抑えるときに不都合

解決策：G-optimal design π^* を用いる

- 適当な x'_t と $\gamma \in [0,1]$ に対して $x_t = (1 - \gamma)x'_t + \gamma\pi^*$ とすると, $S_t^{-1} = S(x_t)^{-1} \preceq \frac{1}{\gamma} S(\pi^*)^{-1}$ より

$$-\hat{\ell}_t^\top a \leq |a_t^\top \ell_t| \cdot |a^\top S_t^{-1} a_t| \leq m \sqrt{|a^\top S_t^{-1} a| \cdot |a_t^\top S_t^{-1} a_t|} \leq \frac{m}{\gamma} G(\pi^*) = \frac{m}{\gamma} N$$

EXP2アルゴリズム

EXP2アルゴリズム

- $\eta > 0$, $\gamma \in [0, 1/2]$, $w_1(a) = 1$ ($\forall a \in A$), $\pi^* \in \Delta(A)$: G-optimal design
- For $t = 1, 2, \dots$
 - $x'_t = \frac{w_t}{\|w_t\|_1}$, $x_t = (1 - \gamma)x'_t + \gamma\pi^*$
 - 分布 x_t に従って a_t を選択, $\ell_t^\top a_t \in [0, m]$ を観測
 - ℓ_t の**普遍推定量**を構成 : $\hat{\ell}_t = S_t^{-1} a_t a_t^\top \ell_t$ ($S_t = \sum_{a \in A} x_t(a) a a^\top$)
 - $w_{t+1}(a) = w_t(a) \exp(-\eta \hat{\ell}_t^\top a)$ ($\forall a \in A$)

定理. $\gamma \geq \eta m N$ のもとで, $R_T := \max_{a^* \in [N]} \mathbf{E}[\sum_{t=1}^T \ell_t^\top a_t - \sum_{t=1}^T \ell_t^\top a^*] \leq \gamma m T + \frac{\log|A|}{\eta} + 2\eta m^2 N T$

したがって $\eta = \Theta\left(\frac{\sqrt{\log N}}{\sqrt{NT}}\right)$, $\gamma = \eta m N$ とおくと $R_T = O(m\sqrt{TN \log|A|})$

EXP2アルゴリズム

- $x'_t = \frac{w_t}{\|w_t\|_1}$, $x_t = (1 - \gamma)x'_t + \gamma\pi^*$, $a_t \sim x_t$
- $\hat{\ell}_t = S_t^{-1} a_t a_t^\top \ell_t$ ($S_t = \sum_{a \in A} x_t(a) a a^\top$), $w_{t+1}(a) = w_t(a) \exp(-\eta \hat{\ell}_t^\top a)$ ($\forall a \in A$)

定理. $\gamma \geq \eta m N$ のもとで, $R_T := \max_{a^* \in [N]} \mathbf{E}[\sum_{t=1}^T \ell_t^\top a_t - \sum_{t=1}^T \ell_t^\top a^*] \leq \gamma m T + \frac{\log|A|}{\eta} + 2\eta m^2 N T$

(証明) x_t, a_t の定義より $\mathbf{E}[\ell_t^\top a_t] = \mathbf{E}[\sum_{a \in A} x_t(a) \ell_t^\top a] \leq \mathbf{E}[\sum_{a \in A} x'_t(a) \ell_t^\top a] + \gamma m$

$$\mathbf{E}[\sum_{t=1}^T \ell_t^\top a_t - \sum_{t=1}^T \ell_t^\top a^*] \leq \mathbf{E}[\sum_{t=1}^T \sum_{a \in A} x'_t(a) \hat{\ell}_t^\top a - \sum_{t=1}^T \hat{\ell}_t^\top a^*] + \gamma m T$$

鏡像降下法のリグレット上界より,

$$\mathbf{E}[\sum_{t=1}^T \sum_{a \in A} x'_t(a) \hat{\ell}_t^\top a - \sum_{t=1}^T \hat{\ell}_t^\top a^*] \leq \frac{\log|A|}{\eta} + \mathbf{E} \left[\sum_{t=1}^T \left(\langle \hat{g}_t, x'_t - x'_{t+1} \rangle - \frac{1}{\eta} D_\psi(x'_{t+1}, x'_t) \right) \right]$$

$$\left(\hat{g}_t \in \mathbb{R}^A, \hat{g}_t(a) = \hat{\ell}_t^\top a \geq -\frac{mN}{\gamma} \geq -\frac{1}{\eta} \right)$$

$$\mathbf{E} \left[\langle \hat{g}_t, x'_t - x'_{t+1} \rangle - \frac{1}{\eta} D_\psi(x'_{t+1}, x'_t) \right] \leq \eta \mathbf{E}[\sum_{a \in A} x'_t(a) \hat{g}_t(a)^2] \leq 2\eta m^2 \mathbf{E}[\sum_{a \in A} x_t(a) (a^\top S_t^{-1} a_t)^2]$$

$$\begin{aligned} \mathbf{E}[\sum_{a \in A} x_t(a) (a^\top S_t^{-1} a_t)^2] &= \mathbf{E}[a_t^\top S_t^{-1} (\sum_{a \in A} x_t(a) a a^\top) S_t^{-1} a_t] = \mathbf{E}[a_t^\top S_t^{-1} S_t S_t^{-1} a_t] \\ &= \text{tr}(\mathbf{E}[S_t^{-1} a_t a_t^\top]) = \text{tr}(S_t^{-1} S_t) = N \end{aligned}$$

問題設定の分類

$$m := \max_{a \in A} \|a\|_1$$

	$A = [N]$	$A \subseteq \{0,1\}^N \cong 2^{[N]}$
完全情報	エキスパート問題 <ul style="list-style-type: none"> • OGD: $R_T = O(\sqrt{TN})$ • Hedge: $R_T = O(\sqrt{T \log N})$ 	完全情報オンライン組合せ最適化 <ul style="list-style-type: none"> • Hedge: $O(\sqrt{m^3 T \log N})$ • OMD: $O(\sqrt{nmT})$ • FTPL: $O(\sqrt{nm^2 T})$
部分観測	多腕バンディット <ul style="list-style-type: none"> • EXP3: $R_T = O(\sqrt{TN \log N})$ 	セミバンディット組合せ最適化 <ul style="list-style-type: none"> • OMD: $O(\sqrt{mTN \log N})$
		バンディット組合せ最適化 <ul style="list-style-type: none"> • EXP2: $O(m\sqrt{TN \log A })$

問題設定の分類

$$m := \max_{a \in A} \|a\|_1$$

	$A = [N]$	$A \subseteq \{0,1\}^N \cong 2^{[N]}$
完全情報	エキスパート問題 <ul style="list-style-type: none"> OGD: $R_T = O(\sqrt{TN})$ Hedge: $R_T = O(\sqrt{T \log N})$ 下界: $R_T = \Omega(\sqrt{T \log N})$ 	完全情報オンライン組合せ最適化 <ul style="list-style-type: none"> Hedge: $O(\sqrt{m^3 T \log N})$ OMD: $O(\sqrt{nmT})$ FTPL: $O(\sqrt{nm^2 T})$
部分観測	多腕バンディット <ul style="list-style-type: none"> EXP3: $R_T = O(\sqrt{TN \log N})$ Tsallis: $R_T = O(\sqrt{TN})$ 下界: $R_T = \Omega(\sqrt{TN})$ 	セミバンディット組合せ最適化 <ul style="list-style-type: none"> OMD: $O(\sqrt{mTN \log N})$ Tsallis: $O(\sqrt{mTN})$
		バンディット組合せ最適化 <ul style="list-style-type: none"> EXP2: $O(m\sqrt{TN \log A })$ Tsallis: $O\left(m\sqrt{TN \log \frac{ A }{N}}\right)$

講義の概要

- リグレット解析の枠組みとその応用
 - オンライン凸最適化問題とオンライン勾配降下法
 - Online-to-batch 変換
 - ミニマックス定理
 - エクスパート問題と Hedge アルゴリズム
- 正則化を用いたアルゴリズムとリグレット解析
 - オンライン鏡像降下法
 - 多腕バンディット問題と EXP3 アルゴリズム
- オンライン組合せ最適化
 - 完全情報フィードバック
 - セミバンディットフィードバック
 - バンディットフィードバック

さらに深く学ぶには

- オンライン凸最適化
 - Hazan, Elad. *Introduction to online convex optimization*. Foundations and Trends in Optimization 2.3-4 (2016): 157-325.
 - Orabona, Francesco. *A modern introduction to online learning*. arXiv preprint arXiv:1912.13213 (2019).
- オンライン学習, learning in games
 - Cesa-Bianchi, Nicolo, and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.
- バンディットアルゴリズム
 - Lattimore, Tor, and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

参考文献

- [AHK12] Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.
- [ACBF02] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47:235–256, 2002.
- [ACBFS02] Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.
- [AK08] Baruch Awerbuch and Robert Kleinberg. Online linear optimization and adaptive routing. *Journal of Computer and System Sciences*, 74(1):97–114, 2008.
- [AK04] Baruch Awerbuch and Robert D Kleinberg. Adaptive routing with end-to-end feedback: Distributed learning and geometric approaches. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 45–53, 2004.
- [CBL06] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.
- [Cov91] Thomas M Cover. Universal portfolios. *Mathematical finance*, 1(1):1–29, 1991.
- [DHS11] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7), 2011.
- [Haz16] Elad Hazan. Introduction to online convex optimization. *Found. Trends Optim.*, 2:157–325, 2016.
- [HAK06] Elad Hazan, Amit Agarwal, and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69:169–192, 2006.
- [HRB07] Elad Hazan, Alexander Rakhlin, and Peter Bartlett. Adaptive online gradient descent. *Advances in Neural Information Processing Systems*, 20, 2007.

参考文献

- [LS20] Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.
- [LW94] Nick Littlestone and Manfred K Warmuth. The weighted majority algorithm. *Information and computation*, 108(2):212–261, 1994.
- [MS10] H Brendan McMahan and Matthew Streeter. Adaptive bound optimization for online convex optimization. *COLT 2010*, page 244, 2010.
- [Ora19] Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019.
- [Vov01] Volodya Vovk. Competitive on-line statistics. *International Statistical Review*, 69(2):213–248, 2001.
- [Zin03] Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *International Conference on Machine Learning*, pages 928–936, 2003.