Budgeted stream-based active learning via adaptive submodular maximization

Kaito Fujii (UTokyo)

(joint work with Hisashi Kashima (KyotoU))

ERATO 感謝祭 Season IV

2017 年 8 月 3 日

目次

1 応用：プール型能動学習とストリーム型能動学習

2 既存研究：適応的劣モジュラ最大化

3 既存研究：劣モジュラ秘書問題

4 提案フレームワーク

5 実験

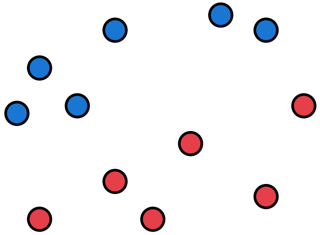
教師あり分類

入力

ラベルつきサンプルの集合 $\{(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}\}_{i=1,\cdots,n}$

出力

ラベルを予測する関数 $\hat{f} : \mathcal{X} \to \mathcal{Y}$



例） $\mathcal{X} = \mathbb{R}^2$,
$\mathcal{Y} = \{\text{red}, \text{blue}\}$

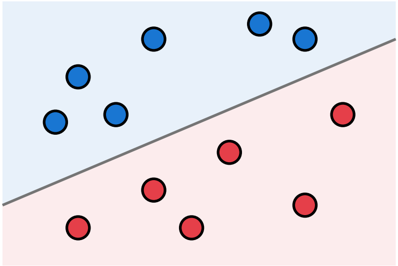
教師あり分類

入力

ラベルつきサンプルの集合 $\{(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}\}_{i=1,\cdots,n}$

出力

ラベルを予測する関数 $\hat{f} : \mathcal{X} \to \mathcal{Y}$



例） $\mathcal{X} = \mathbb{R}^2$,
$\mathcal{Y} = \{\mathrm{red}, \mathrm{blue}\}$

能動学習の動機

現実には

- ラベルなしサンプルはたくさん手に入るが，
- ラベル付与にはお金や時間がかかる

という状況が多い

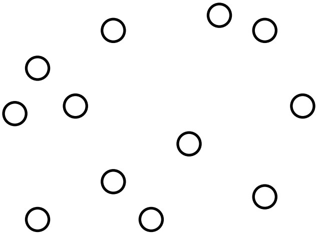
⇓

能動学習

ラベルを付与するサンプルを**能動的に**選ぶことで，
少ないラベルつきサンプルで高い予測精度

プール型能動学習

ラベルなしサンプルは最初にすべて与えられる



ラベルなしサンプルの集合

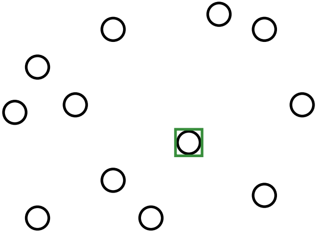
$V = \{x_i\}_{i=1,\dots,n} \subset \mathcal{X}$

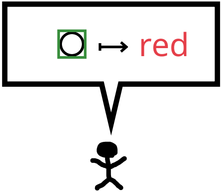
ラベル付与オラクル

$\phi : V \to \mathcal{Y}$

プール型能動学習

ラベルなしサンプルは最初にすべて与えられる





ラベルなしサンプルの集合

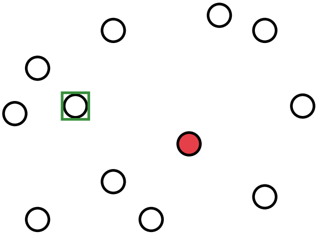
$V = \{x_i\}_{i=1,\cdots,n} \subset \mathcal{X}$

ラベル付与オラクル

$\phi : V \to \mathcal{Y}$

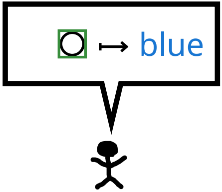
プール型能動学習

ラベルなしサンプルは最初にすべて与えられる



ラベルなしサンプルの集合

$V = \{x_i\}_{i=1,\cdots,n} \subset \mathcal{X}$

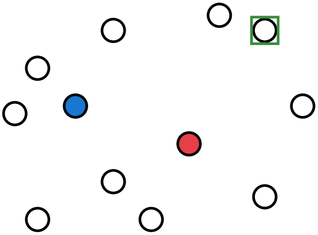


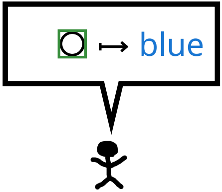
ラベル付与オラクル

$\phi : V \to \mathcal{Y}$

プール型能動学習

ラベルなしサンプルは最初にすべて与えられる





ラベルなしサンプルの集合

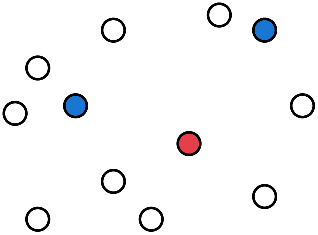
$V = \{x_i\}_{i=1,\dots,n} \subset \mathcal{X}$

ラベル付与オラクル

$\phi : V \to \mathcal{Y}$

プール型能動学習

ラベルなしサンプルは最初にすべて与えられる



ラベルなしサンプルの集合

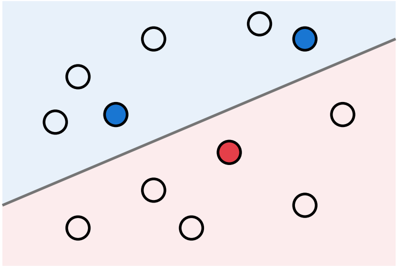
$V = \{x_i\}_{i=1,\dots,n} \subset \mathcal{X}$

ラベル付与オラクル

$\phi: V \to \mathcal{Y}$

プール型能動学習

ラベルなしサンプルは最初にすべて与えられる



ラベルなしサンプルの集合

$V = \{x_i\}_{i=1,\cdots,n} \subset \mathcal{X}$

ラベル付与オラクル

$\phi : V \to \mathcal{Y}$

ストリーム型能動学習

ラベルなしサンプルは一つずつ現れる

n 個のうち k 個にラベル付与する設定を考える（n, k 既知）

ラベルなしサンプルの集合
$V = \{x_i\}_{i=1,\cdots,n} \subset \mathcal{X}$

ラベル付与オラクル
$\phi: V \to \mathcal{Y}$

ストリーム型能動学習

ラベルなしサンプルは一つずつ現れる

n 個のうち k 個にラベル付与する設定を考える（n, k 既知）

ラベルなしサンプルの集合

$V = \{x_i\}_{i=1,\cdots,n} \subset \mathcal{X}$

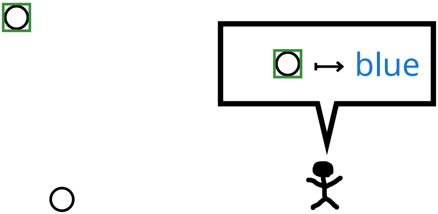
ラベル付与オラクル

$\phi : V \to \mathcal{Y}$

ストリーム型能動学習

ラベルなしサンプルは一つずつ現れる

n 個のうち k 個にラベル付与する設定を考える（n, k 既知）



ラベルなしサンプルの集合

$$V = \{x_i\}_{i=1,\cdots,n} \subset \mathcal{X}$$

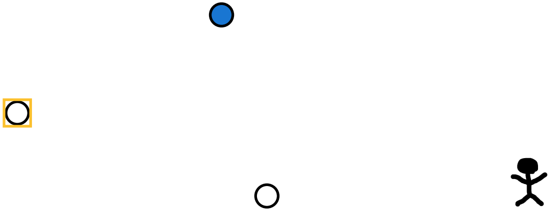
ラベル付与オラクル

$$\phi : V \to \mathcal{Y}$$

ストリーム型能動学習

ラベルなしサンプルは一つずつ現れる

n 個のうち k 個にラベル付与する設定を考える（n, k 既知）



ラベルなしサンプルの集合

$$V = \{x_i\}_{i=1,\cdots,n} \subset \mathcal{X}$$

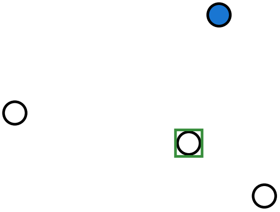
ラベル付与オラクル

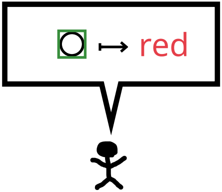
$$\phi \colon V \to \mathcal{Y}$$

ストリーム型能動学習

ラベルなしサンプルは一つずつ現れる

n 個のうち k 個にラベル付与する設定を考える（n, k 既知）





ラベルなしサンプルの集合

$V = \{x_i\}_{i=1,\cdots,n} \subset \mathcal{X}$

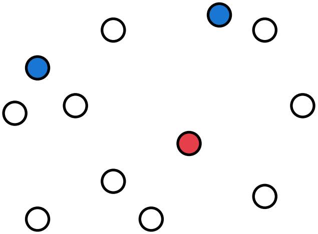
ラベル付与オラクル

$\phi \colon V \to \mathcal{Y}$

ストリーム型能動学習

ラベルなしサンプルは一つずつ現れる

n 個のうち k 個にラベル付与する設定を考える（n, k 既知）



ラベルなしサンプルの集合

$V = \{x_i\}_{i=1,\cdots,n} \subset \mathcal{X}$

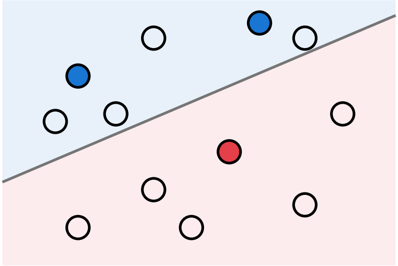
ラベル付与オラクル

$\phi : V \to \mathcal{Y}$

ストリーム型能動学習

ラベルなしサンプルは一つずつ現れる

n 個のうち k 個にラベル付与する設定を考える（n, k 既知）



ラベルなしサンプルの集合

$V = \{x_i\}_{i=1,\cdots,n} \subset \mathcal{X}$

ラベル付与オラクル

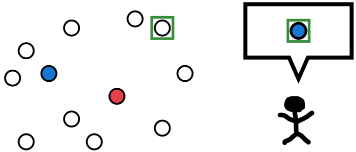
$\phi : V \to \mathcal{Y}$

本研究の概要

ストリーム型能動学習のための組合せ最適化フレームワーク

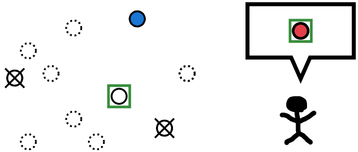
プール型能動学習

ラベルなしサンプルはすべて既知



ストリーム型能動学習

サンプルはひとつずつ現れる



適応的劣モジュラ最大化
[Golovin–Krause'11]

→

新たなフレームワーク
（本研究）

↑

劣モジュラ秘書問題
[Bateni–Hajiaghayi–Zadimoghaddam'13]

目次

1 応用：プール型能動学習とストリーム型能動学習

2 既存研究：適応的劣モジュラ最大化

3 既存研究：劣モジュラ秘書問題

4 提案フレームワーク

5 実験

劣モジュラ最大化

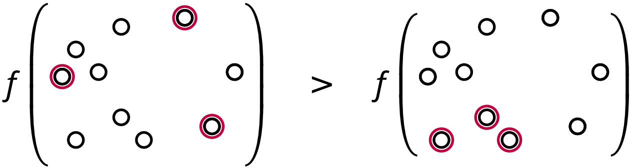
与えられた集合 V の “よい” 部分集合を探す問題

$$
\begin{aligned}
&\text{Maximize} \quad && f(S) \\
&\text{subject to} \quad && |S| \le k
\end{aligned}
$$

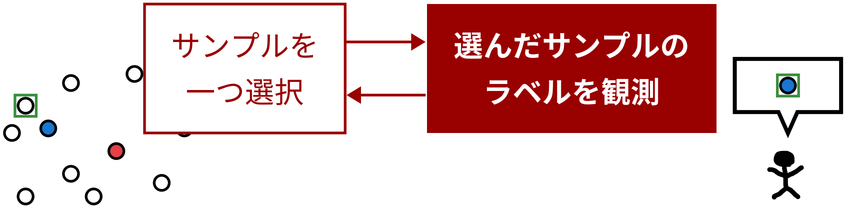
$f : 2^V \to \mathbb{R}$

単調劣モジュラ関数

例）データセット V を要約する部分集合を探す問題



適応的な設定



観測したラベルによって次に選ぶサンプルを変えられる

→ 同じアルゴリズムでも，目的関数値が変化

適応的劣モジュラ最大化 [Golovin–Krause'11]

オラクルに関する事前分布を考え，目的関数の期待値を最大化

劣モジュラ性と単調性

$f(v|S) \triangleq f(S \cup \{v\}) - f(S)$

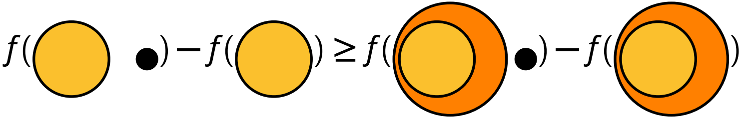
$S \subseteq V$ に対して $v \in V$ を追加したときの**増分**

劣モジュラ性と単調性

$f(v|S) \triangleq f(S \cup \{v\}) - f(S)$

$S \subseteq V$ に対して $v \in V$ を追加したときの**増分**

劣モジュラ性 $f(v|A) \geq f(v|B)$ $(\forall A \subseteq B, \forall v \in V \setminus B)$

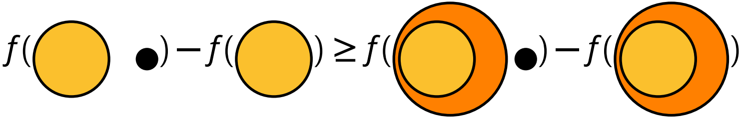


劣モジュラ性と単調性

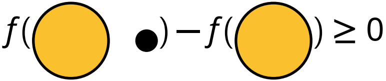
$f(v|S) \triangleq f(S \cup \{v\}) - f(S)$

$S \subseteq V$ に対して $v \in V$ を追加したときの**増分**

劣モジュラ性 $f(v|A) \geq f(v|B)$ $(\forall A \subseteq B, \forall v \in V \setminus B)$



単調性 $f(v|A) \geq 0$ $(\forall A \subseteq V, \forall v \in V)$



適応的劣モジュラ性と適応的単調性 [Golovin–Krause'11]

$\Delta(v|\psi)$

オラクルから $\psi = \{(v_i, y_i)\}_{i=1,\dots,\ell}$ を観測しているとき，
新たに $v \in V$ を追加した場合の**増分の期待値**

適応的劣モジュラ性

$$\Delta(v|\psi_1) \geq \Delta(v|\psi_2) \ (\forall \psi_1 \subseteq \psi_2, \forall v \in V \setminus \mathrm{dom}(\psi_2))$$

適応的単調性

$$\Delta(v|\psi) \geq 0 \ (\forall \psi, \forall v \in V)$$

目次

1 応用：プール型能動学習とストリーム型能動学習

2 既存研究：適応的劣モジュラ最大化

3 既存研究：劣モジュラ秘書問題

4 提案フレームワーク

5 実験

古典的秘書問題 [folklore'60s]

問題

ランダムな順序で n 人の候補を面接する（n は既知）

面接したら直ちに採用するかどうか決める

アルゴリズム

最初の $\lfloor n/e \rfloor$ 人は採用しない

それ以降にその時点までで最良の候補が現れたら採用

➡ 1/e 以上の確率で最良の候補を採用できる

古典的秘書問題 [folklore'60s]

問題

ランダムな順序で n 人の候補を面接する（n は既知）

面接したら直ちに採用するかどうか決める

アルゴリズム

最初の $\lfloor n/e \rfloor$ 人は採用しない

それ以降にその時点までで最良の候補が現れたら採用

➡ 1/e 以上の確率で最良の候補を採用できる

$n = 10$

3

古典的秘書問題 [folklore’60s]

問題

ランダムな順序で n 人の候補を面接する（n は既知）

面接したら直ちに採用するかどうか決める

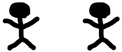
アルゴリズム

最初の $\lfloor n/e \rfloor$ 人は採用しない

それ以降にその時点までで最良の候補が現れたら採用

→ 1/e 以上の確率で最良の候補を採用できる

$n = 10$



3 4

古典的秘書問題 [folklore’60s]

問題

ランダムな順序で n 人の候補を面接する（n は既知）

面接したら直ちに採用するかどうか決める

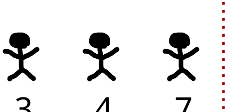
アルゴリズム

最初の $\lfloor n/e \rfloor$ 人は採用しない

それ以降にその時点までで最良の候補が現れたら採用

⟶ 1/e 以上の確率で最良の候補を採用できる

$n = 10$



3 4 7

古典的秘書問題 [folklore'60s]

問題

ランダムな順序で n 人の候補を面接する（n は既知）

面接したら直ちに採用するかどうか決める

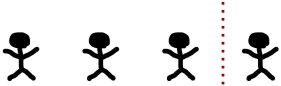
アルゴリズム

最初の $\lfloor n/e \rfloor$ 人は採用しない

それ以降にその時点までで最良の候補が現れたら採用

➡ 1/e 以上の確率で最良の候補を採用できる

$n = 10$



3　4　7　6

古典的秘書問題 [folklore’60s]

問題

ランダムな順序で n 人の候補を面接する（n は既知）

面接したら直ちに採用するかどうか決める

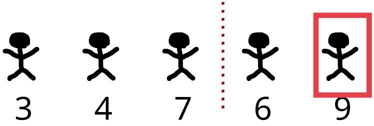
アルゴリズム

最初の $\lfloor n/e \rfloor$ 人は採用しない

それ以降にその時点までで最良の候補が現れたら採用

➡ 1/e 以上の確率で最良の候補を採用できる

$n = 10$



古典的秘書問題 [folklore'60s]

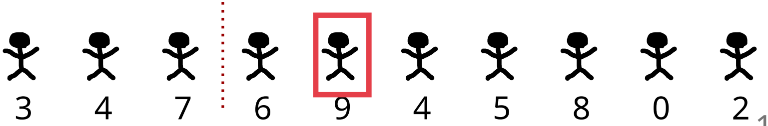
問題

ランダムな順序で n 人の候補を面接する（n は既知）
面接したら直ちに採用するかどうか決める

アルゴリズム

最初の $\lfloor n/e \rfloor$ 人は採用しない
それ以降にその時点までで最良の候補が現れたら採用
→ 1/e 以上の確率で最良の候補を採用できる

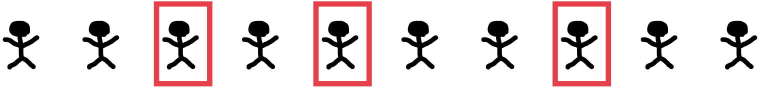
$n = 10$



劣モジュラ秘書問題 [Bateni–Hajiaghayi–Zadimoghaddam'13]

ランダムな順序で現れる n 人の候補のなかから k 人を選ぶ
単調劣モジュラ関数 $f\colon 2^V \to \mathbb{R}$ を最大化

$n = 10, k = 3$



競合比

$\alpha \in [0, 1]$ がアルゴリズムの競合比

$\overset{\text{def}}{\Longleftrightarrow}$

任意の問題例について，アルゴリズムの解 S が

$$\mathbb{E}[f(S)] \geq \alpha \underbrace{\max_{S^* \subseteq V} f(S^*)}_{\text{入力をすべて知っている場合の最適値}}$$

を満たす

目次

1 応用：プール型能動学習とストリーム型能動学習

2 既存研究：適応的劣モジュラ最大化

3 既存研究：劣モジュラ秘書問題

4 提案フレームワーク

5 実験

提案フレームワークの仮定

仮定 1　ストリームの順序に**ランダム性**を仮定

仮定 2　全体のサンプル数 n は既知だと仮定

仮定 3　目的関数に**方策適応的劣モジュラ性**を仮定

- 適応的劣モジュラ性よりも強い性質
- 応用で使われる多くの関数で成り立っている

適応的劣モジュラ性との比較

劣モジュラ性

$$f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B)$$
$$(\forall A \subseteq B, v \in V \setminus B)$$

適応的設定に拡張

適応的劣モジュラ性

$$\Delta(v|\psi_A) \geq \Delta(v|\psi_B)$$
$$(\forall \psi_A \subseteq \psi_B, v \in V \setminus B)$$

適応的劣モジュラ性との比較

劣モジュラ性

$$f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B) \quad (\forall A \subseteq B, v \in V \setminus B)$$

$$f(A \cup P) - f(A) \geq f(B \cup P) - f(B) \quad (\forall A \subseteq B, P \subseteq V \setminus B)$$

適応的設定に拡張

適応的劣モジュラ性

$$\Delta(v|\psi_A) \geq \Delta(v|\psi_B) \quad (\forall \psi_A \subseteq \psi_B, v \in V \setminus B)$$

適応的劣モジュラ性との比較

劣モジュラ性

$$f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B) \quad (\forall A \subseteq B,\, v \in V \setminus B)$$

$$\longleftarrow \qquad \longrightarrow$$

$$f(A \cup P) - f(A) \geq f(B \cup P) - f(B) \quad (\forall A \subseteq B,\, P \subseteq V \setminus B)$$

適応的設定に拡張

適応的劣モジュラ性

$$\Delta(v | \psi_A) \geq \Delta(v | \psi_B) \quad (\forall \psi_A \subseteq \psi_B,\, v \in V \setminus B)$$

$$\longleftarrow \qquad \not\longrightarrow$$

方策適応的劣モジュラ性

$$\Delta(\pi | \psi_A) \geq \Delta(\pi | \psi_B) \quad (\forall \psi_A \subseteq \psi_B,\, \pi(V) \subseteq V \setminus B)$$

方策適応的劣モジュラ関数の性質

補題 [Feige–Mirrokni–Vondrák'07]

$f : 2^V \to \mathbb{R}$ 劣モジュラ関数.

$A(r)$: $A \subseteq V$ の各要素を確率 r で含む集合.

$$\mathbb{E}[f(A(r))] \geq (1-r)f(\emptyset) + rf(A)$$

方策適応的劣モジュラ関数の性質

補題 [Feige–Mirrokni–Vondrák'07]

$f : 2^V \to \mathbb{R}$ 劣モジュラ関数.

$A(r)$: $A \subseteq V$ の各要素を確率 r で含む集合.

$$\mathbb{E}[f(A(r))] \geq (1-r)f(\emptyset) + rf(A)$$

適応的設定に拡張

補題 [Fujii–Kashima'16]

$(\{f_\phi\}_\phi, p)$ 方策適応的劣モジュラ関数.

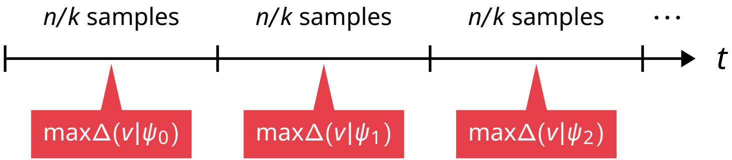
$\pi(r)$: 方策 π から各要素を確率 $1-r$ で削除してできる方策.

$$\mathbb{E}[f_\phi(A(\pi(r), \phi))] \geq (1-r)\mathbb{E}[f_\phi(\emptyset)] + r\mathbb{E}[f_\phi(A(\pi, \phi))]$$

適応的ストリームアルゴリズム

ストリーム設定 限られたメモリの範囲でサンプルを保持

ストリーム全体を k 分割し，
各部分ストリームから増分の期待値最大のサンプルを選択



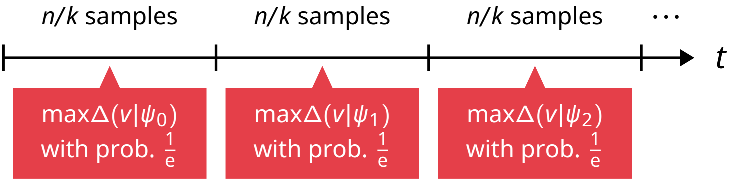
定理 [Fujii–Kashima'16]

このアルゴリズムの競合比は $(2-\sqrt{3})(1-1/e) \approx 0.16$.

適応的秘書アルゴリズム

秘書設定 ラベルを付与するかどうかすぐに判断

各部分ストリームに古典的秘書アルゴリズムを適用して，
増分の期待値最大のサンプルを確率 1/e で選ぶ



定理 [Fujii–Kashima'16]

このアルゴリズムの競合比は $\dfrac{1-1/e}{2e\sqrt{1+2/e}} \approx 0.08.$

目次

1 応用：プール型能動学習とストリーム型能動学習
2 既存研究：適応的劣モジュラ最大化
3 既存研究：劣モジュラ秘書問題
4 提案フレームワーク
5 実験

実験設定

データセット

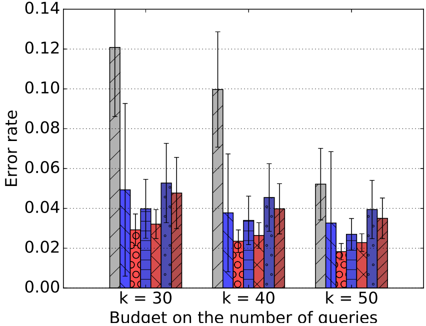
- WDBC（$n = 596$，32 次元）
- MNIST（$n = 14780$，PCA で 10 次元に削減）

比較手法

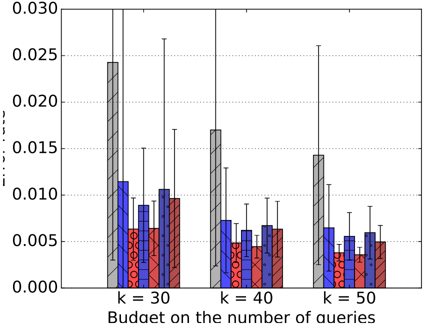
- 不確実性サンプリング
- ランダムサンプリング

提案手法は ALuMA [Gonen+13] に基づいて設計

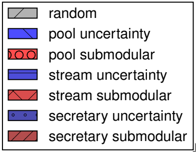
実験結果



WDBC



MNIST



それぞれの設定で

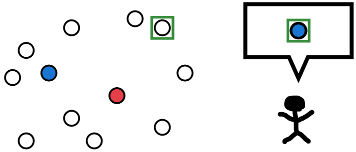
提案手法は比較手法よりも誤差が小さい

本研究の概要

ストリーム型能動学習のための組合せ最適化フレームワーク

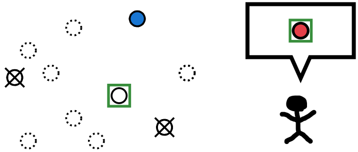
プール型能動学習

ラベルなしサンプルはすべて既知



ストリーム型能動学習

サンプルはひとつずつ現れる



適応的劣モジュラ最大化
[Golovin–Krause’11]

→

新たなフレームワーク
（本研究）

↑

劣モジュラ秘書問題
[Bateni–Hajiaghayi–Zadimoghaddam’13]