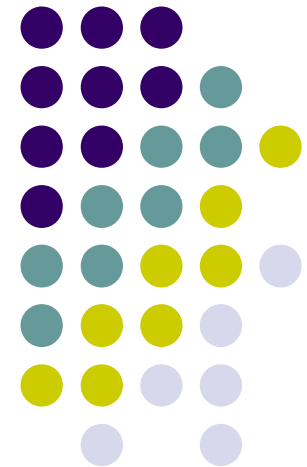


ICML 2007 を読む会 Discriminative Learning for Differing Training and Test Distributions

S. Bickel, M. Bruckner, T. Scheffer

読む人: 坪井祐太
(日本IBM 東京基礎研究所)



論文の概要

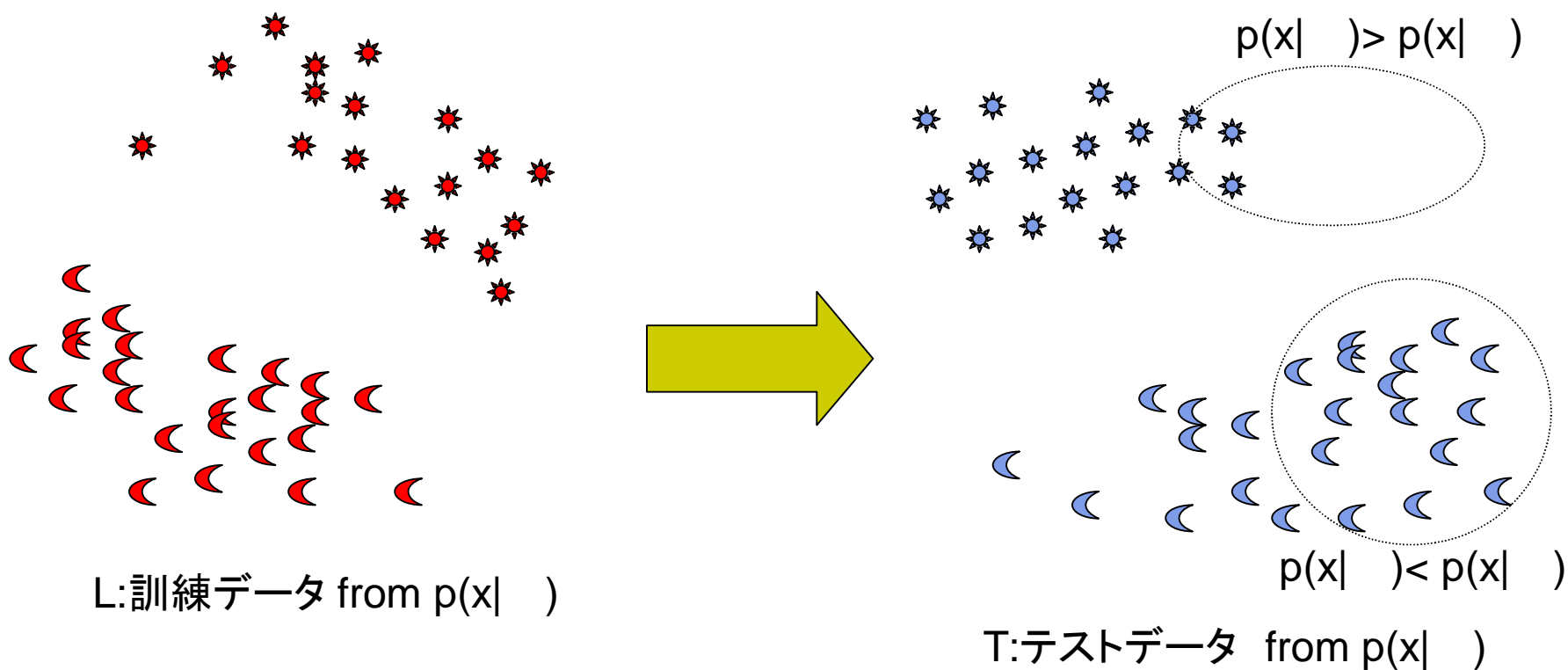
Discriminative Learning for Differing Training and Test Distributions



- 問題: 訓練データとテストデータの入力分布が異なる(共変量シフト)場合の分類
- 解法: ラベル付き訓練データ L とラベルなしテストデータ T 进行分类する識別問題
- 特徴: 重要度の推定と分類器の学習を同時に行う

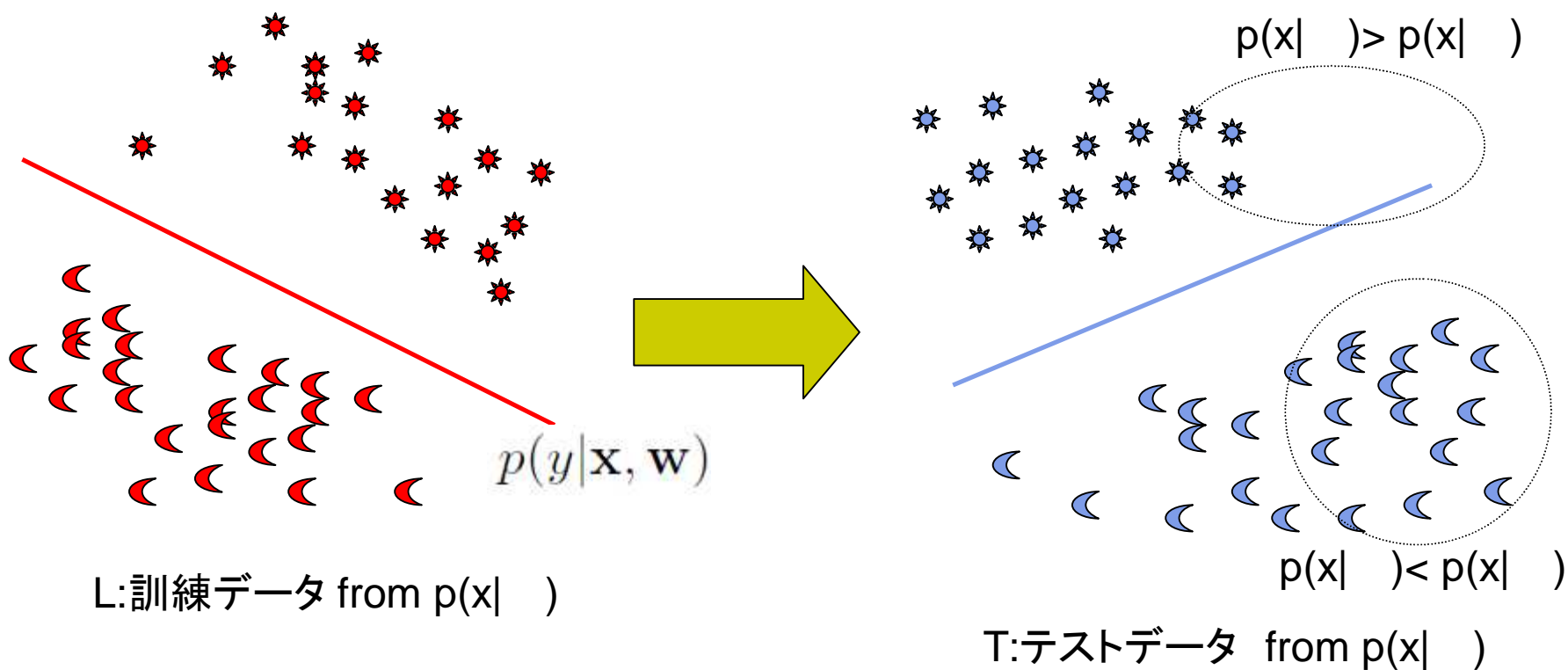
Classification under Covariate shift

訓練データとテストデータの入力分布 $p(x)$ が異なる



Classification under Covariate shift

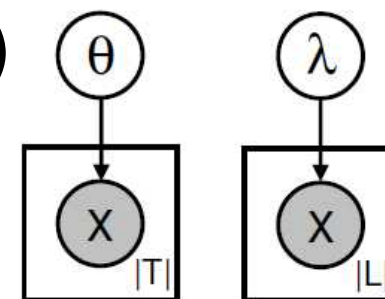
訓練データとテストデータの入力分布 $p(x)$ が異なる





Covariate shift先行研究

- L: ラベル付き訓練データ from $p(\mathbf{x}|\theta)$
- T: テストデータ from $p(\mathbf{x}|\lambda)$
- Shimodaira 2000



$$\begin{aligned} & p(y|\mathbf{x}) \text{ が変わらないならば、} \\ & \int p(\mathbf{x}, y|\theta) \text{Loss}(f(\mathbf{x}), y) d\mathbf{x} \\ &= \int p(\mathbf{x}, y|\lambda) \frac{p(\mathbf{x}|\theta)}{p(\mathbf{x}|\lambda)} \text{Loss}(f(\mathbf{x}), y) d\mathbf{x} \end{aligned}$$

テスト分布での期待損失
が重み付けされた訓練分
布での期待損失と等しい

- Huang et al. (2007): Kernel Mean Matching

訓練データ \mathbf{x} の各 $\frac{p(\mathbf{x}|\theta)}{p(\mathbf{x}|\lambda)}$ を直接推定

Discriminative Learning for Differing Distributions

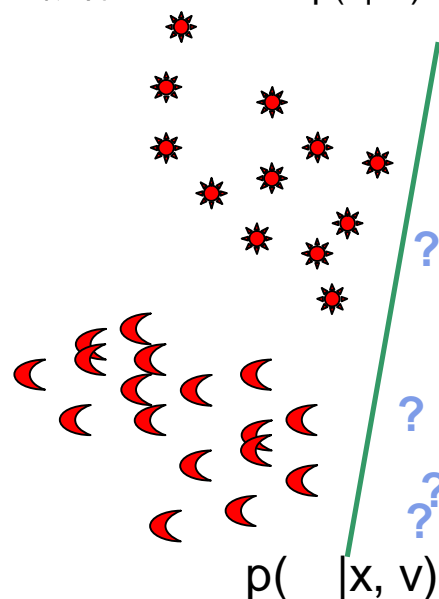
$\sigma=1$: 訓練データ、 $\sigma=0$: テストデータ を示す変数
条件付確率 $p(\sigma | \mathbf{x}, \mathbf{v})$ を識別学習 (\mathbf{v} : 学習パラメータ)



$$\frac{p(\mathbf{x}|\theta)}{p(\mathbf{x}|\lambda)} = \frac{p(\sigma=1|\theta, \lambda)}{p(\sigma=0|\theta, \lambda)} \frac{p(\sigma=0|\theta, \lambda)}{p(\sigma=1|\theta, \lambda)} \frac{p(\mathbf{x}|\theta)}{p(\mathbf{x}|\lambda)}$$

$$= \frac{p(\sigma=1|\theta, \lambda)}{p(\sigma=0|\theta, \lambda)} \left(\frac{1}{p(\sigma=1|\mathbf{x}, \theta, \lambda)} - 1 \right)$$

L: 訓練データ from $p(\mathbf{x}|\lambda)$



T: テストデータ from $p(\mathbf{x}|\lambda)$

最尤推定値は m/n
 m : 訓練データ数
 n : テストデータ数

$p(\sigma=1 | \mathbf{x}, \mathbf{v})$ とし、 \mathbf{v} を学習

$$\frac{1}{p(\sigma=1 | \mathbf{x}, \mathbf{v})} - 1 = \frac{p(\sigma=0 | \mathbf{x}, \mathbf{v})}{p(\sigma=1 | \mathbf{x}, \mathbf{v})}$$

Discriminative Learning for Differing Distributions

$p(\mathbf{x}|\mathbf{v})$ と分類器 $p(y|\mathbf{x}, \mathbf{w})$ のパラメータ \mathbf{w} と \mathbf{v} の事後分布 $p(\mathbf{w}, \mathbf{v}|L, T)$ を同時にMAP推定



$$\begin{aligned}
 & p(\mathbf{w}, \mathbf{v}|L, T) \quad \text{尤度} \quad \text{事前分布} \\
 & \propto P(L|\mathbf{w}, \mathbf{v}) P(L, T|\mathbf{v}) p(\mathbf{w}) p(\mathbf{v}) \\
 & = \left(\prod_{i=1}^m P(y_i|\mathbf{x}_i; \mathbf{w}) \frac{p(\sigma=1|\mathbf{v})}{p(\sigma=0|\mathbf{v})} \left(\frac{1}{p(\sigma_i=1|\mathbf{x}_i; \mathbf{v})} - 1 \right) \right) \\
 & \quad \left(\prod_{i=1}^{m+n} P(\sigma_i|\mathbf{x}_i; \mathbf{v}) \right) p(\mathbf{w}) p(\mathbf{v}) \quad \text{重み付き尤度 for 分類器 } p(y|\mathbf{x}, \mathbf{w}) \\
 & \quad \text{尤度 for } p(\mathbf{x}|\mathbf{v})
 \end{aligned}$$

Kernel Logistic Regressionによるモデル化

$p(\sigma = 1 | \mathbf{x}, \mathbf{v})$ と分類器 $p(y | \mathbf{x}, \mathbf{w})$ を(カーネル)ロジスティック回帰でモデリング



- 条件付確率

$$p(y = 1 | \mathbf{x}; \mathbf{w}) = \frac{1}{1 + \exp(-\mathbf{w}^T \mathbf{x})}$$

$$p(\sigma = 1 | \mathbf{x}; \mathbf{v}) = \frac{1}{1 + \exp(-\mathbf{v}^T \mathbf{x})}$$

- 事前分布

$$p(\mathbf{w}) \propto \exp\left(\frac{-\mathbf{w}^T \mathbf{w}}{2s_w^2}\right); \quad p(\mathbf{v}) \propto \exp\left(\frac{-\mathbf{v}^T \mathbf{v}}{2s_v^2}\right)$$

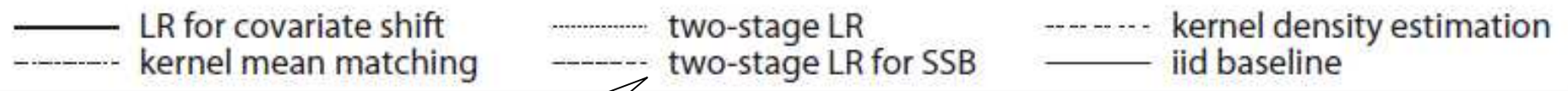
- ニュートン法で最適化(非凸最適化)

実験結果1(spam filtering)

test: 9人のmail box 32-2048 mails, training: 1000 mails

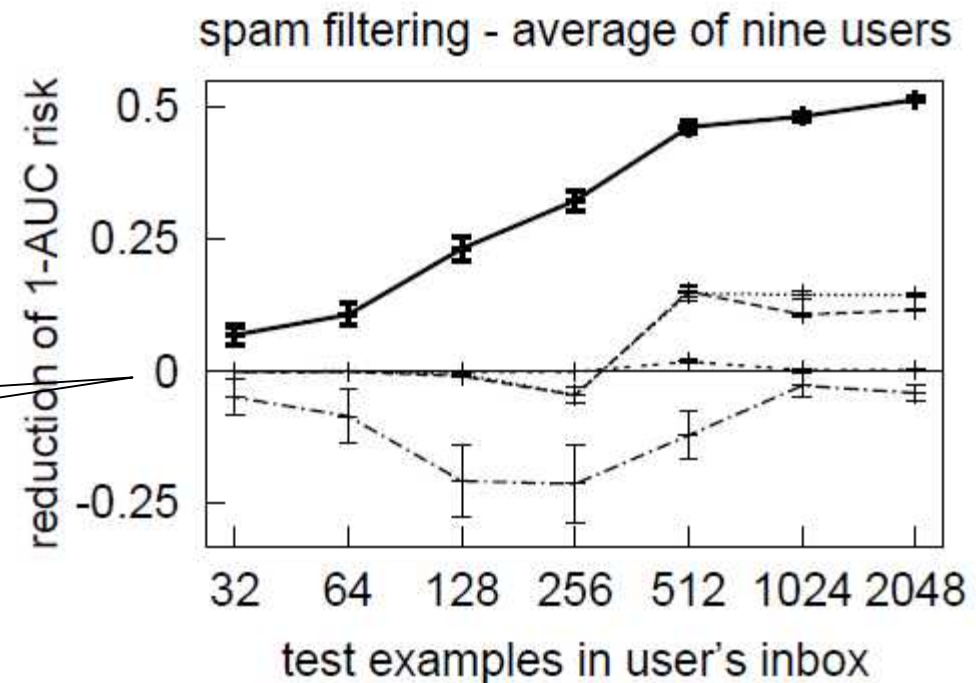
linear kernel (834,661次元)

評価: 1-AUC (area under the ROC curve) riskの削減率



$p(s=1, |x, v)^{-1}$ で重み付け

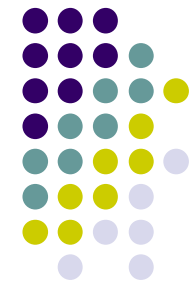
Baseline AUC=0.992



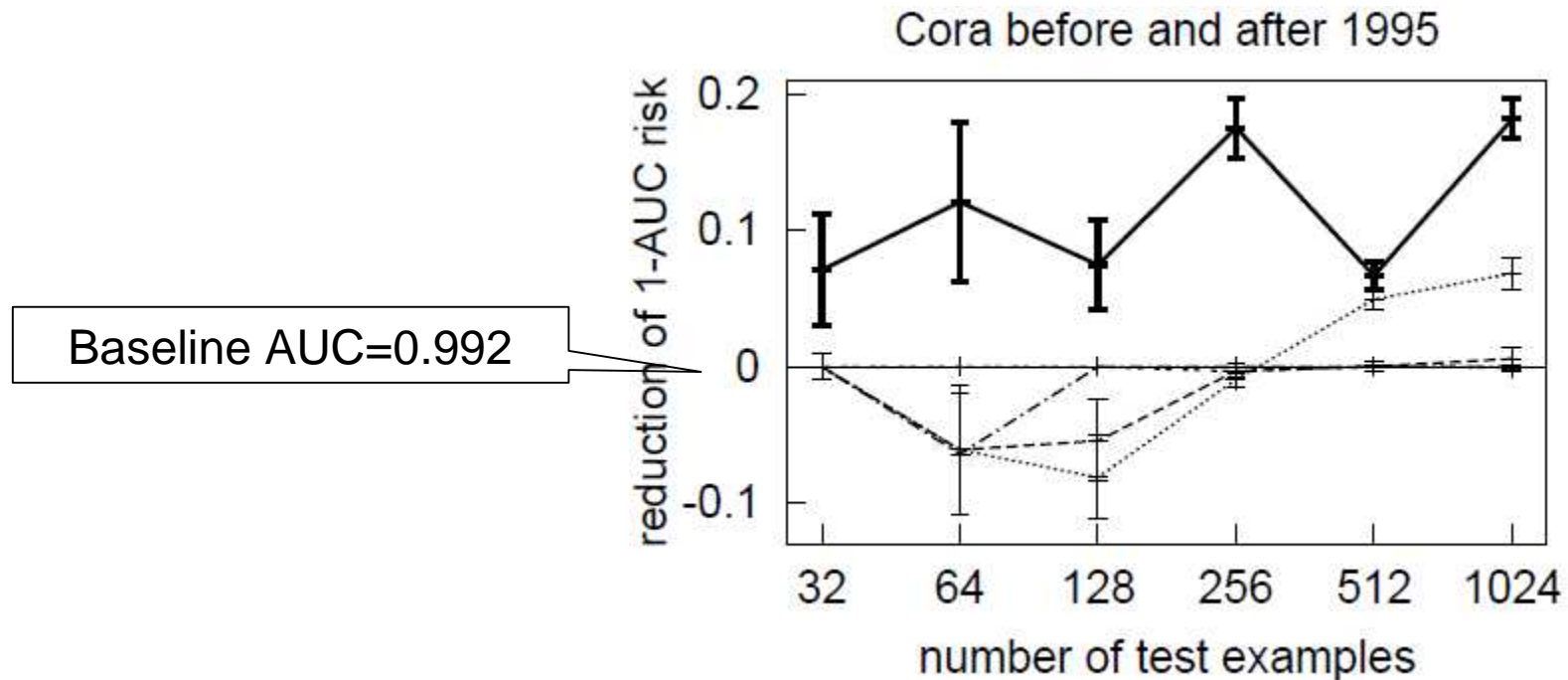
実験結果2 (論文分類: Machine LearningとNetworking)

test: 1995年以降 32-1024, training: 1995年以前 1219

linear kernel (40,000次元: 単語)



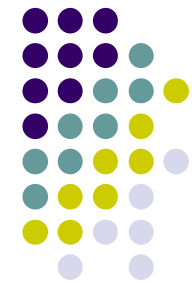
— LR for covariate shift two-stage LR	----- kernel density estimation
- - - - - kernel mean matching	- - - - - two-stage LR for SSB	—— iid baseline



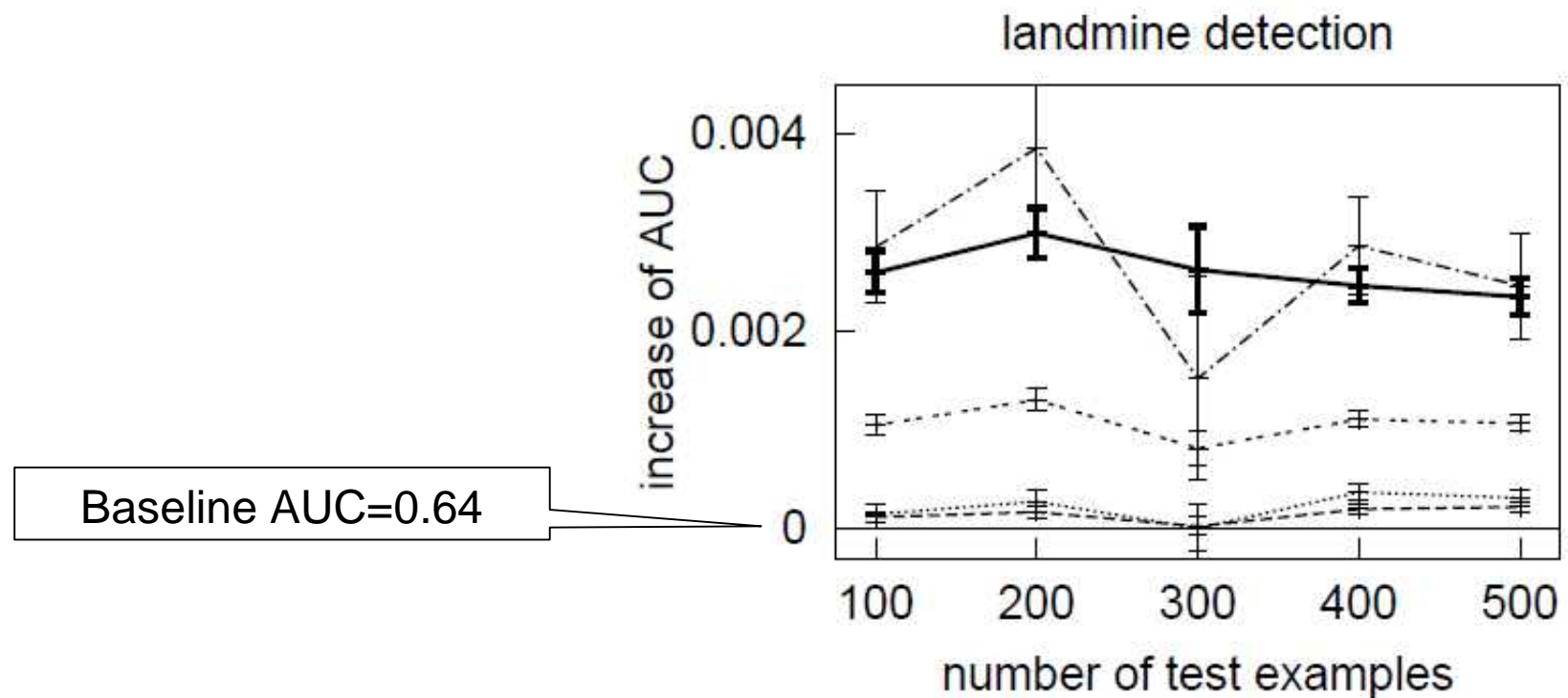
実験結果3 (地雷検出)

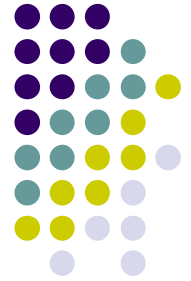
test:地雷原100-500, training: 別の地雷原 500

RBF kernel (9次元:レーダー情報)



— LR for covariate shift two-stage LR	----- kernel density estimation
- - - - - kernel mean matching	- - - - - two-stage LR for SSB	—— iid baseline





まとめ

- 訓練データとテストデータの入力分布が異なる時の分類器のMAP学習を提案
- ラベル付き訓練データ L とラベルなしテストデータ T を識別する分類器と、ラベル付与する分類器の学習を同時に行う