



<第4回>

統計モデリングと機械学習

統計数理研究所

松井 知子

tmatsui@ism.ac.jp

目次

1. 機械学習とは
2. 機械学習の応用先
3. 機会学習の一般的な問題
4. カーネル法
5. 協調フィルタリング
6. まとめ
7. 参考文献

1. 機械学習とは

- データからの帰納
- McKinsey Global Instituteのレポート(2011)
「機械学習はイノベーションの次の大きな波を呼ぶ」
- P. Domingos, “A few useful things to know about machine learning,” Comm. of The ACM, 55(10), 2012
「機械学習によるアプリケーションを成功させるには、いくつかの“**folk knowledge**”を知っておく必要がある」

1. 機械学習とは

- 機械学習～

統計的

コーディング + モデリング + 目的関数 + 最適化

1. 機械学習とは

- 機械学習～

統計的

コーディング + モデリング + 目的関数 + 最適化

アプリケーションドメイン
の専門家

1. 機械学習とは

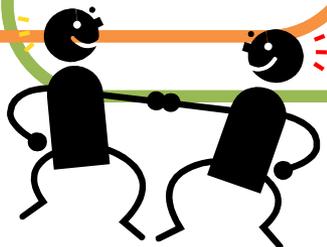
- 機械学習～

統計的

コーディング + モデリング + 目的関数 + 最適化

アプリケーションドメイン
の専門家

機械学習の専門家



1. 機械学習とは

- モデリング
 - 線形回帰/識別モデル、ニューラルネットワーク、ロジスティック回帰、グラフィカルモデル、混合ガウス分布、隠れマルコフモデル、決定木 など
- 目的関数
 - 最小二乗誤差、最小KLダイバージェンス、最大マージン、最尤、最大事後確率 など
- 最適化
 - 連続最適化: 最急降下法、共役勾配法、ニュートン法、EMアルゴリズム など
 - 離散最適化: ビーム探索、greedy探索 など

2. 機械学習の応用先

いろいろな分類・回帰予測:

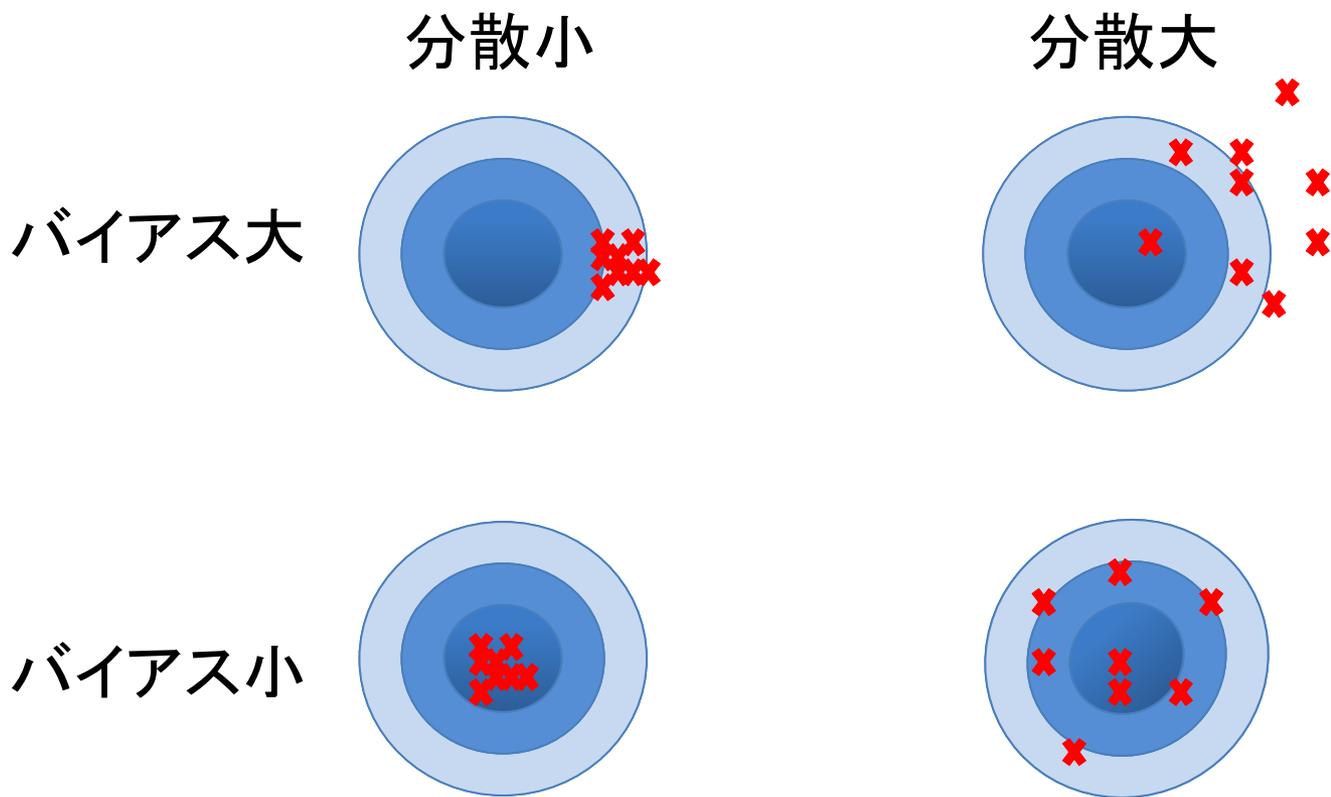
- 画像・音声・話者認識などのパターン認識
 - 情報検索
 - 金融市場の予測
 - DNA配列の分類
 - ロボディクス
- など

3. 機械学習の一般的な問題

- No-free-lunch定理 [Wolpert and Macready, 1995]
「どの学習機械に対しても、ランダム推測の方が優れている
目的関数が少なくとも一つ存在する」
- みにくいアヒルの子定理 [渡辺慧, 1969]
「課題から独立した特徴量／モデルは存在しない」

3. 機械学習の一般的な問題

- バイアスと分散



3. 機械学習の一般的な問題

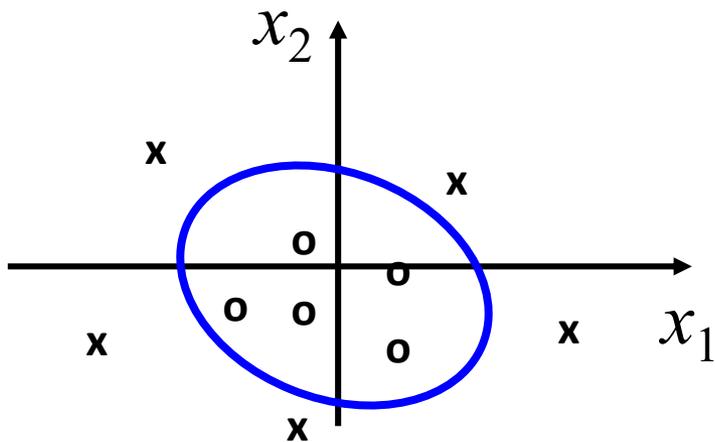
- データ量の増大
 - 高性能計算 (HPC)
 - グラフィックス プロセッシング ユニット (GPU)
 - クーダ (CUDA) などの専用の開発環境
 - 高速探索アルゴリズム
- データ量の不足
 - ➡ 協調フィルタリング など

3. 機械学習の一般的な問題

- 次元の呪い[Bellman, 1961]
 - 空間の次元が増えるのに対応して、問題の算法が指数関数的に大きくなる。
 - 高次元空間では直感が効きにくい？

➡カーネル法 など

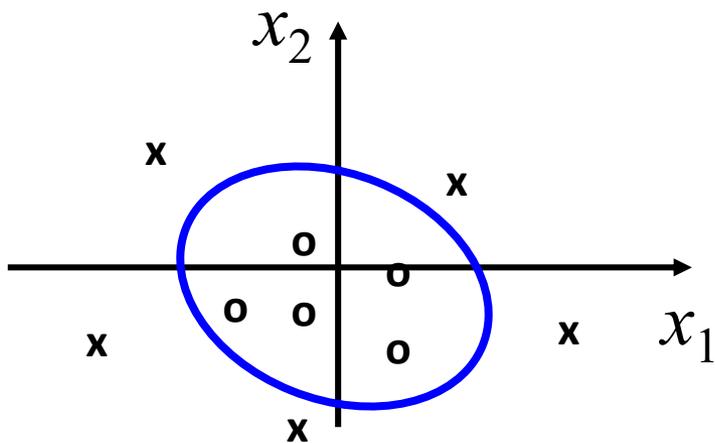
4. カーネル法



(x_1, x_2)

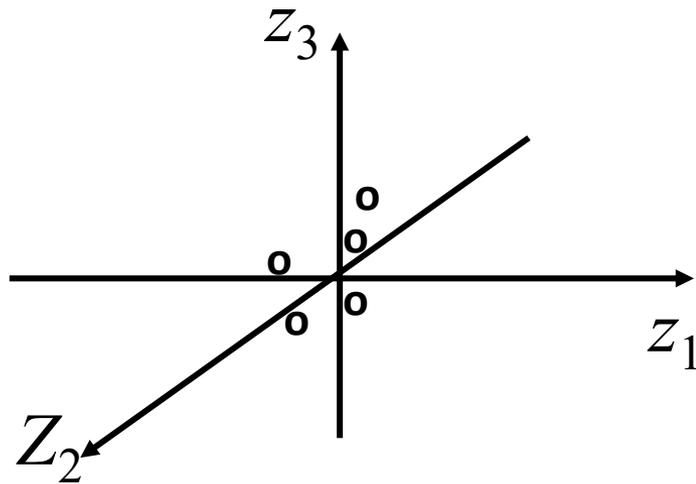
入力空間 R^2

4. カーネル法



$$(x_1, x_2)$$

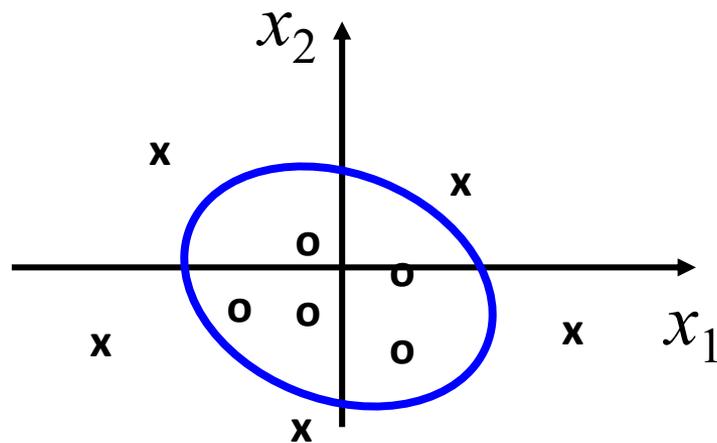
入力空間 R^2



$$(z_1, z_2, z_3) := (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$

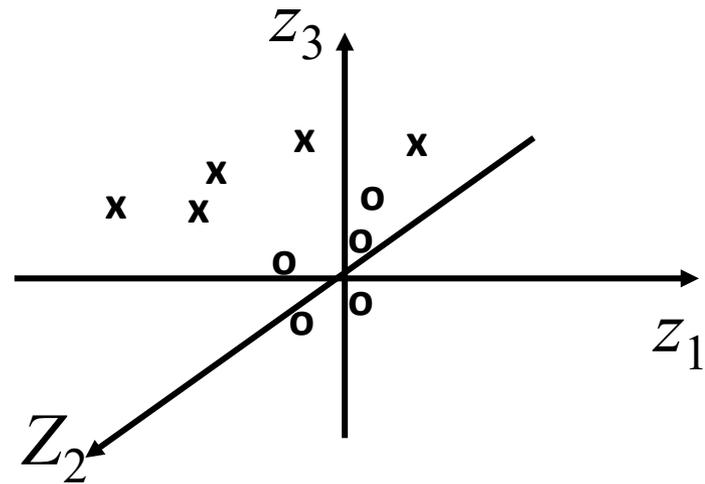
特徴空間 R^3

4. カーネル法



$$(x_1, x_2)$$

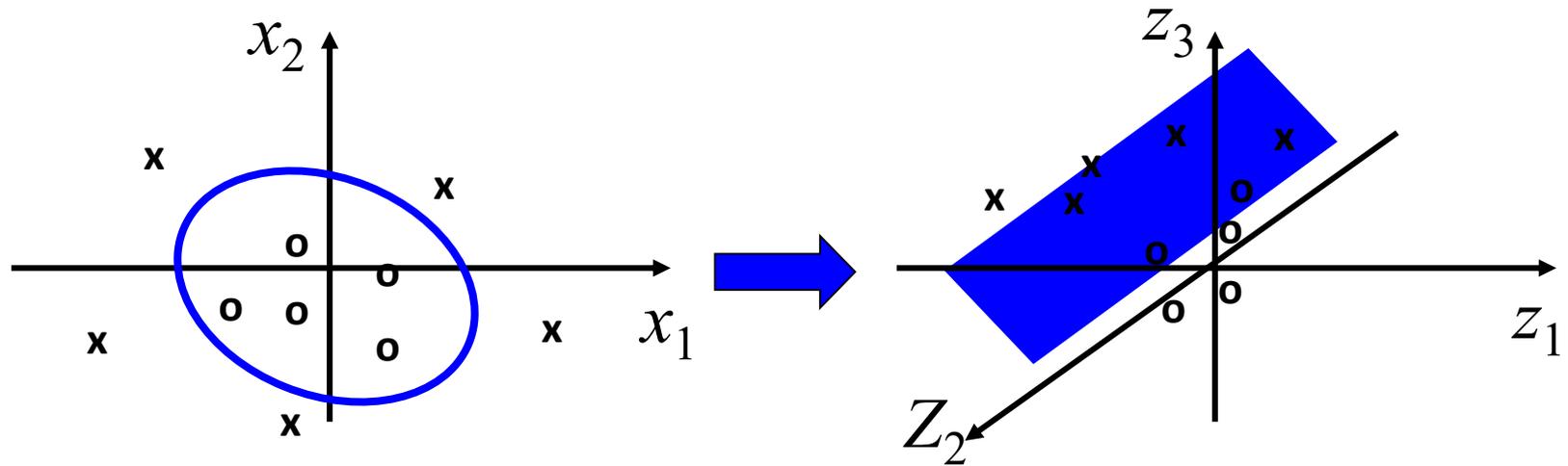
入力空間 R^2



$$(z_1, z_2, z_3) := (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$

特徴空間 R^3

4. カーネル法



$$(x_1, x_2)$$

$$(z_1, z_2, z_3) := (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$

入力空間 R^2

特徴空間 R^3

4. カーネル法

- 2次元から3次元空間への写像

$$\mathbf{x} = (x_1, x_2)$$

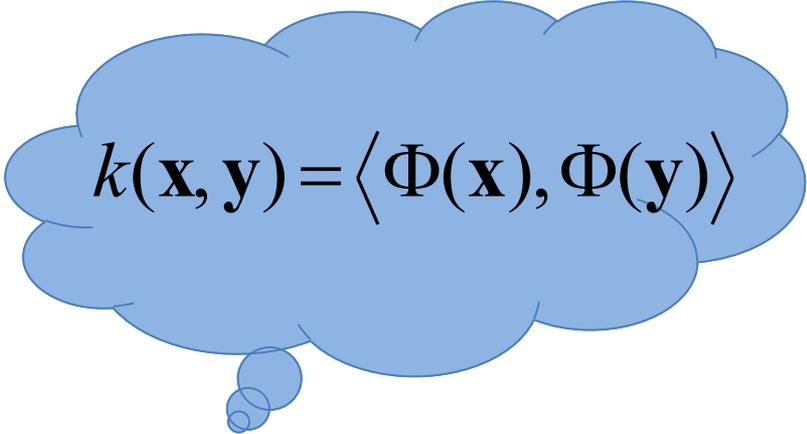
$$\Phi(\mathbf{x}) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$

4. カーネル法

- 2次元から3次元空間への写像

$$\mathbf{x} = (x_1, x_2)$$

$$\Phi(\mathbf{x}) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$


$$k(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle$$

- カーネル関数

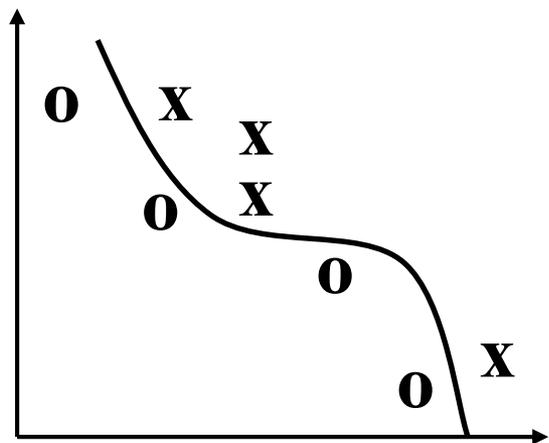
$$k(\mathbf{x}, \mathbf{y}) := \langle \mathbf{x}, \mathbf{y} \rangle^2$$

$$\langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle = \langle (x_1^2, \sqrt{2}x_1x_2, x_2^2), (y_1^2, \sqrt{2}y_1y_2, y_2^2) \rangle$$

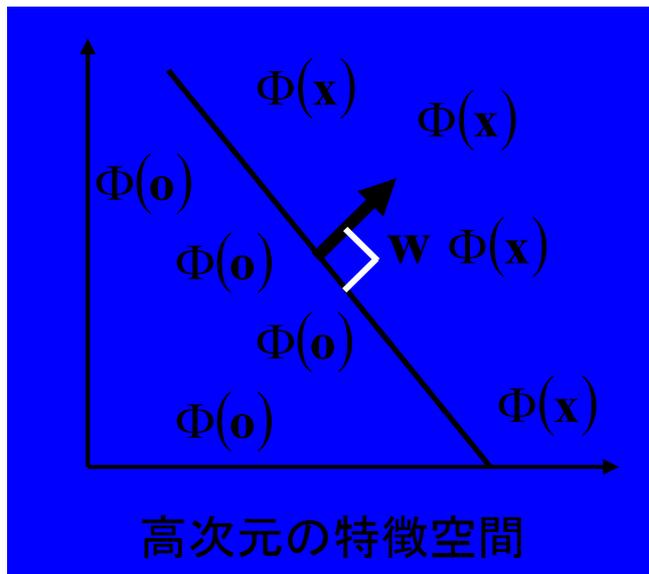
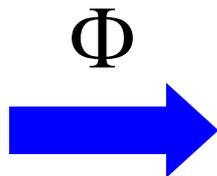
$$= \langle (x_1, x_2), (y_1, y_2) \rangle^2$$

$$= \langle \mathbf{x}, \mathbf{y} \rangle^2$$

4. カーネル法



入力空間



高次元の特徴空間

入力空間では線形分離不能

識別関数

$$f(x) = \mathbf{w}^T \Phi(x) + b$$

$$f(x) \geq 0 : \text{class } \mathbf{x}$$

$$f(x) < 0 : \text{class } \mathbf{o}$$

5. 協調フィルタリング

	映画1	映画2	映画3	..
ユーザ1		5		..
ユーザ2		3		..
ユーザ3	5		1	..
:	:	:	:	..

- ユーザID: p_u 映画ID: q_i 評価: r_{ui}

$$\hat{r}_{ui} = \hat{q}_i^T \hat{p}_u$$

$$O(Q, P) = \frac{1}{2} \sum_{(u,i) \in \Omega} (r_{ui} - q_i^T p_u)^2 + \lambda (\|q_i\|^2 + \|p_u\|^2)$$

5. 協調フィルタリング

- Netflix Prize (2006-2009)
映画のユーザレイトィング予測に関するコンペ
 - 優勝賞金 : US\$1,000,000
 - 学習データ : 100,480,507 ratings
(480,189 users、17,770 movies)
 - 優勝者 : BellKor's Pragmatic Chaos team
 - 数百の協調フィルタリングによる学習機械のアンサンブル

6. まとめ

- 統計的モデリングによる機械学習について紹介した。
 - 応用先
 - 一般的な問題
 - カーネル法
 - 協調フィルタリング

7. 参考文献

1. 『パターン認識と機械学習』 C.M. ビショップ (著), 元田 浩 (監訳), 栗田 多喜夫 (監訳), 樋口 知之 (監訳), 松本 裕治 (監訳), 村田 昇 (監訳), 丸善出版
2. 『カーネル法入門—正定値カーネルによるデータ解析』 福水健次 (著), 朝倉書店
3. 『カーネル多変量解析—非線形データ解析の新しい展開』 赤穂 昭太郎 (著), 岩波書店
4. Netflix Prize, <http://www.netflixprize.com/index>
5. Y. Koren, R. Bell and C. Volinsky, “Matrix Factorization Techniques for Recommender Systems,” IEEE Computer, Vol. 42 , 8, pp. 30 – 37, 2009.