# 第8章　サポートベクターマシン

8.1　パーセプトロンの限界
8.2　線形SVM
8.3非線形SVM
8.4.5分析例

03T4027N　佐々　知広

# 線形分離可能(不可能)

## 線形分離可能

ー線形関数による超平面(2次元の場合は直線)で
対象の分類が可能である状況

例、データ配置が論理積(AND)、論理積の否定(NAND)、
和の否定(NOR)といった論理演算の幾何学的表現の集合

## 線形分離不可能

ー線形関数による超平面で対象の分類が不可能である状況

例、データ配置が排他的論理和(XOR)の幾何学的表現の集合

# パーセプトロンの限界

- パーセプトロンの限界
  - パーセプトロンでは線形分離可能な問題でしか解決を保証できない事

- 非線形分離不可能問題（XOR型）への対処

  隠れユニットを利用する

  ＝入力ユニットの情報をシグモイド関数や
  ラジアル既定関数を利用して高次元空間に写像する。

結果：高次元において線形分離可能になり超平面が構成可能

# サポートベクターマシン

- 教師あり学習を用いる識別手法の一つ
- 未学習データに対して高い識別性能を得る為の工夫がある。
- 各データ点との距離が最大となる超平面を求める
マージン最大化という基準で学習する。
- 学習過程はラグランジュ乗法数を用いることにより、
最適化問題の一種である凸二次計画問題で定式化される。

# 線形SVM

$$y_i = \begin{cases} 1 & f(x_i) \geq 0 \text{ のとき} \\ -1 & f(x_i) < 0 \text{ のとき} \end{cases}$$

$$f(x_i) = \omega' x_i + b$$

$y_i$ :オブザベーションiの二値型の名義的反応

$x_i$ : $y_i$ を予測するのに用いるオブザベーションiに関する j 次元のベクトル。

$\omega$ : j次元の重みベクトル　　$b$ :閾値（スカラー）

二つのクラスを区別する超平面は次のように表現

$$f(x) = \omega' x + b = 0$$

# マージン最大化

- マージン

  :二つの識別境界から等距離の地点に実線が描画される。
  識別境界から実線までの距離をマージンと呼ぶ。

データ $x_i$ と超平面との最小距離

$$\underset{i=1,2,\cdots,I}{Min}\left(\frac{|\omega' x_i + b|}{\|\omega\|}\right) \quad (8.4) \qquad \underset{i=1,2,\cdots,I}{Min}(\omega' x_i + b) = 1 \quad (8.5)$$

識別超平面を挟む2つの境界面

$$f(x) = \omega' x_i + b = \pm 1 \quad (8.6)$$

(8.5)より(8.4)は $1/\|\omega\|$ と表現できる。

マージンは2つ定義されるから実際は $2/\|\omega\|$

# 双対問題とラグランジュ未定乗数法

目的関数: $\dfrac{\|\omega\|^2}{2}$　不等式制約: $y_i(\omega' x_i + b) \geq 1$　(8.7)

関数の最大化を関数の2乗の最小化で達成する。

不等式制約はマージン内にデータが存在しないことを表現

ラグランジュ関数 $L(\omega, b, \alpha) = \dfrac{1}{2}\|\omega\|^2 - \sum_{i=1}^{I} \alpha_i [y_i\{(\omega' x_i + b) - 1\}]$

(8.8)

$\alpha_i$ :ラグランジュの未定乗数と呼ばれる母数 $\alpha_i \geq 0$

(8.8)を最小化するωとbにおいて(8.8)に対する接線の傾きは0

$$\frac{\partial L(\omega, b, \alpha)}{\partial \omega} = \omega - \sum_{i=1}^{I} \alpha_i y_i x_i = 0 \Leftrightarrow \omega = \sum_{i=1}^{I} \alpha_i y_i x_i$$

$$\frac{\partial L(\omega, b, \alpha)}{\partial b} = \sum_{i=1}^{I} y_i \alpha_i = 0$$

# 双対問題とラグランジュ未定乗数法

前頁の結果を(8.8)式に代入する。

$$L(\alpha)=\sum_{i=1}^{I}\alpha_i-\frac{1}{2}\sum_{i=1}^{I}\sum_{i*=1}^{I}\alpha_i\alpha_i y_i y_i x_i^{'} x_i \qquad (8.11)$$

制約条件: $\alpha_i \geq 0 \quad \sum_{i=1}^{I} a_i y_i = 0$

$$b=\frac{1}{2}\omega^{'}X_{SV1}+\omega^{'}X_{SV2}$$

$X_{SV1}, X_{SV2}$ はそれぞれクラス1クラス2に所属するオブザベーションのうち$\alpha_i \neq 0$ 満たすものである。サポートベクターという。

(8.11)式は主問題に対する双対問題という。

# 双対問題の利点

- ラグランジュ未定乗数法における不等式制約よりもより簡単な制約化で最適化が行われる。

- 超平面は$\alpha_i \neq 0$ に対応するサポートベクター$X_{SV}$の情報のみによって決定されるので極めて効率のよい計算が可能。

- ωがXの内積$(X'X)$ で定義されているため後に説明する
 カーネルトリックの結果をそのまま代入できる。

# 非線形SVM

- 線形分離不可能問題に対しては高次元空間への写像が効果的である。

高次元空間への写像 $\Phi(x_i)$ を考慮し次の双対問題を最大化する。

$$L(\alpha)=\sum_{i=1}^{I}\alpha_i-\frac{1}{2}\sum_{i=1}^{I}\sum_{i*=1}^{I}\alpha_i\alpha_i y_i y_i\Phi(x_i)'\Phi(x_{i*}) \tag{8.13}$$

カーネルトリック

:$\Phi(x_i)$の内積を $x_i$ の内積の関数として表現したものをカーネル( $k(x_i' x_{i*})$ )といい、カーネルを利用することで写像関数$\Phi(x_i)$ の計算を回避する事が出来る。

# さまざまなカーネル

線形カーネル: $k(x_i^{'} x_j) = x_i^{'} x_j$

多項式カーネル: $k(x_i^{'} x_j) = (\gamma x_i^{'} x_j + \delta)^d$

RBF(ガウネシアン)カーネル: $k(x_i^{'} x_j) = \exp(\gamma \| x_i - x_j \|^2)$

シグモイドカーネル: $k(x_i^{'} x_j) = \tanh(\gamma (x_i\, x_j) - \delta)$

上記に存在する未知のパラメータは識別器の精度に
大きく影響するため、モデルの交差妥当性に配慮しつつ、
適当な値にチューニングする必要がある。

# ソフトマージン

- 少数のオブザベーションの存在によって線形分離不可能となる場合に、非線形SVMを適用することはかならずしも好ましくない。

ソフトマージン

:少数のオブザベーションの誤分類をゆるすマージン最大化学習則

- 次の不等式制約を考える。

$$y_i(\omega' x_i + b) \geq 1 - \xi_i, \xi_i \geq 0$$

$\xi$:スラック変数

C:ペナルティ母数境界面からの逸脱度を伝える母数

目的関数:$\dfrac{\|\omega\|^2}{2} + C = \sum_{i=1}^{N} \xi_i, C > 0$

# モデルのチューニング

- グリッドサーチ

  :一定の間隔をもった数値を配しヒューリスティックに
  適切なチューニング母数を探索する手法。
  一般にまずきめの粗いグリッドによって最適な母数の
  検討をつけその後きめの細かいグリッドによって
  最適母数を決定する。

- 交差妥当化

  :チューニング母数が未知のデータに対して
  妥当なものとなるように行う手法。
  学習データをk群に分割し、そのうちの一群を検証データ、
  残りのk-1群を学習データとして全k回交差妥当化検証を行う。