

線形 SVM

1 はじめに

サポートベクターマシン (svm) とは、教師あり学習の 1 種である。サポートベクターマシンの基本的な考え方はマージン最大化に基づく超平面を用いて与えられたデータセットに対してクラス分類を行うことである。サポートベクターマシンは 2 クラス分類に対してもっとも効果的である。また、線形分類モデルのサポートベクターマシンとこれを非線形モデルにまで拡張したサポートベクターマシンが存在するが、ここでは前者を扱うものとする。

2 マージン最大化分類

導入では、サポートベクターマシンの基本的な考え方は超平面を用いて与えられたデータを分類することだと述べた。ここではその具体的な操作を見ていくものとする。

話を具体化するために 2 次元におけるクラス分類を用いて説明する。以下の図のデータセットは線分によってクラス分類できることは直感的にわかる。その際に、線分の方程式は以下のようなとなる。

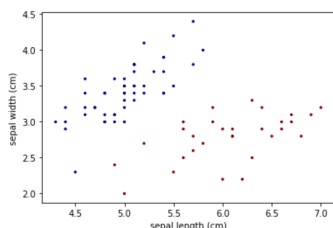


図 1: 直線により分類できるデータセット

$$w_1 x_1 + w_2 x_2 + c = 0 \quad (1)$$

この式はベクトル表記を用いて以下のように表現できる。

$$\mathbf{w}^T \mathbf{x} + \gamma = 0 \quad (2)$$

ただし、 $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $\gamma \in \mathbb{R}$

この線分より上に存在する点 \mathbf{x}_+ について $\mathbf{w}^T \mathbf{x}_+ + \gamma > 0$ 、下に存在する点 \mathbf{x}_- について $\mathbf{w}^T \mathbf{x}_- + \gamma < 0$ が成立することがわかる。

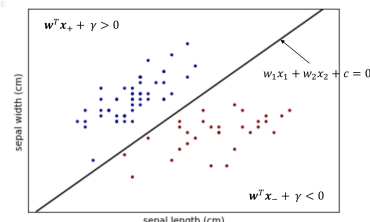


図 2: 境界線とサンプルの関係

ここで、サポートベクターマシンは教師あり学習であるので、データセット $D = \{\mathbf{x}_i, y_i\}$ について、 $y_i = +1$ または $y_i = -1$ というラベルが予め与えられているとすると、データ・セットの各点について以下の式が成立すれば正しいクラス分類ができたと言える。

$$(\mathbf{w}_i^T \mathbf{x}_i + \gamma) y_i > 0 \quad (3)$$

次に、マージンを定義することとする。マージンを m_i とし、

$$m_i = (\mathbf{w}_i^T \mathbf{x}_i + \gamma) y_i \quad (4)$$

と置く。マージンはデータセットが直線 $\mathbf{w}^T \mathbf{x} + \gamma$ と最も距離が近い点の間隔のようなものと考えることができる。また、上述のような点は「サポートベクター」と呼ばれる。以後の計算はサポートベクターが直線 $\mathbf{w}^T \mathbf{x} + \gamma = 1$ 上にあると仮定して計算をしていくこととする。

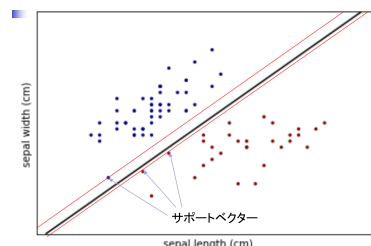


図 3: サポートベクター

サポートベクターが直線 $(\mathbf{w}^T \mathbf{x} + \gamma) = 1$ 上にあるとき、サポートベクターと直線 $(\mathbf{w}^T \mathbf{x} + \gamma) = 0$ 間の距離は以下の式によって計算される。

$$\frac{(\mathbf{w}^T \mathbf{x}_i + \gamma) y_i}{\|\mathbf{w}\|_2} \quad (5)$$

サポートベクターが $w^T x_i + \gamma = 1$ 上にあるとき、サポートベクターを含む各クラスのデータは $w^T x_i + \gamma \geq 1$ における領域にある。故に、

$$\min \frac{(w^T x_i + \gamma)y_i}{\|w\|_2} = \frac{1}{\|w\|_2} \quad (6)$$

$\frac{1}{\|w\|_2}$ はこのサポートベクターマシンのマージンの最小値である。しかし、サポートベクターマシンのマージンの最小値を最大化することにより、データ変動に強いサポートベクターマシンを作成できる。故に、

$$\max \frac{1}{\|w\|_2} \quad s.t. \quad w^T x_i + \gamma \geq 1 \quad (7)$$

を求めれば最大マージンを計算できるが、この問題は以下の問題に同置である。

$$\min \|w\|_2 \quad s.t. \quad w^T x_i + \gamma \geq 1 \quad (8)$$

以上に仮定したサポートベクターマシンはサポートベクターを境界として必ず超平面を用いて与えられたサンプルを正しく分類できると仮定した。このようなものをハードマージン・サポートベクターマシンという。しかし、現実では与えられたサンプルは超平面によりきれいに分けることができないものが多い。その際に、マージンについての制限 $(w^T x_i + \gamma)y_i \geq 1$ を緩めたものをソフトマージン・サポートベクターマシンという。

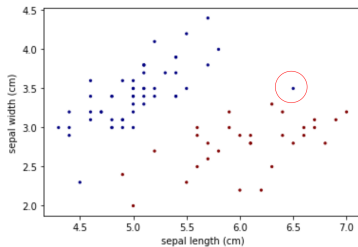


図 4: 超平面によりサンプルを分類できない例

3 サポートベクトル分類器の解

この節では第2節にて述べたソフトマージン・サポートベクターマシンの解について扱う。ソフトマージン・サポートベクターマシンではマージンについての制限を緩和し、以下のようとなる。

$$(w^T x_i + \gamma) \geq 1 - \xi_i \quad (9)$$

ただし、 $\xi_i \geq 0, \forall i = 0, 1, \dots, n$

つまり、サポートベクターが定めた超平面を破るようなサンプルがあってもそれを認めることで制限緩和を実現し

ている。制限緩和することにより、マージンを最大化する問題は以下のように書き換えられる。

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad s.t. \quad (w^T x_i + \gamma)y_i \geq 1 - \xi_i, \xi_i \geq 0 \quad (10)$$

この問題の解を求めるためにラグランジュ未定乗数法を用いる。その際のラグランジュ関数を以下のように置く。

$$L(w, \gamma, \xi, \alpha, \beta) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i \{(w^T x_i + \gamma)y_i - 1 + \xi_i\} - \sum_{i=1}^n \beta_i \xi_i \quad (11)$$

故に求めたい問題は、

$$\min_{w, \gamma, \xi} \max_{\alpha, \beta} L(w, \gamma, \xi, \alpha, \beta) \quad s.t. \quad \alpha \geq 0, \beta \geq 0 \quad (12)$$

に書き換えられる。 $\min_{w, \gamma, \xi} L(w, \gamma, \xi, \alpha, \beta)$ の最適性条件は、

$$\frac{\partial L}{\partial w} = 0 \quad (13)$$

$$\frac{\partial L}{\partial \gamma} = 0 \quad (14)$$

$$\frac{\partial L}{\partial \xi} = 0 \quad (15)$$

すなわち、

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (16)$$

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (17)$$

$$\alpha + \beta = C \quad (18)$$

この結果により、 w は α によって書き換えられ、 $\alpha + \beta = C$ により ξ と γ により消去できることが分かる。したがって元の問題は以上の結果を用いると以下のように書き換えられる。

$$\max_{\alpha} L(\alpha) \quad s.t. \quad 0 \leq \alpha \leq C, \sum_{i=1}^n \alpha_i y_i = 0 \quad (19)$$

ただし、 $L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j$
故に、

$$\hat{\alpha} = \arg \max_{\alpha} L(\alpha) \quad s.t. \quad 0 \leq \alpha \leq C, \sum_{i=1}^n \alpha_i y_i = 0 \quad (20)$$

以上のような $\hat{\alpha}$ を求めることができれば、そこから \hat{w} と $\hat{\gamma}$ を算出できる。 $\hat{\alpha}$ についての最適化問題は二次計画問題に分類される。二次計画問題については <http://www.support-vector-machines.org/> にて公開されているソフトウェアを用いて解くことができる。なお、 $\hat{\gamma}$ の求め方については次節にて紹介する。

4 スパース性

第5章では ℓ_1 制限のもとでは解がスパースになることを学習した。しかし、サポートベクターマシンにおいては双対解 $\hat{\alpha}$ は ℓ_1 制限がなくともスパースになる傾向があることを解説する。

$\min_{\mathbf{w}, \gamma, \xi} \max_{\alpha, \beta} L(\mathbf{w}, \gamma, \xi, \alpha, \beta) \quad s.t. \quad \alpha \geq 0, \beta \geq 0$ は KKT 最適性条件を満たすので、

$$\alpha_i(m_i - 1 + \xi_i) = 0 \quad (21)$$

$$\beta_i \xi_i = 0 \quad (22)$$

これに $\alpha + \beta = C$ を代入すると、以下の結果が得られる。

$$\alpha_i = 0 \implies m_i > 1 \quad (23)$$

$$0 < \alpha_i < C \implies m_i = 1 \quad (24)$$

$$\alpha_i = C \implies m_i < 1 \quad (25)$$

$$m_i > 1 \implies \alpha_i = 0 \quad (26)$$

$$m_i = 1 \implies 0 \leq \alpha_i \leq C \quad (27)$$

$$m_i < 1 \implies \alpha_i = C \quad (28)$$

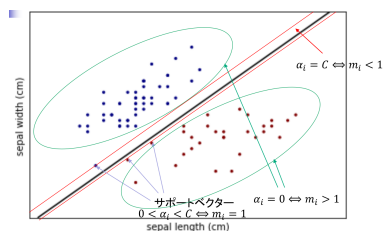


図 5: m_i と α_i の関係

特に、式 24 の条件におけるサンプルを「サポートベクトル」と呼ぶ。このとき、 $m_i = (\mathbf{w}^T \mathbf{x}_i + \gamma)y_i = 1$ となるので、

$$\begin{aligned} \hat{\gamma} &= \frac{1}{y_i} - \hat{\mathbf{w}}^T \mathbf{x}_i \\ &= y_i - \hat{\mathbf{w}}^T \mathbf{x}_i \\ &= y_i - \sum_{j=1}^n \hat{\alpha}_j y_j \mathbf{x}_j^T \mathbf{x}_i \\ &= y_i - \sum_{j=1}^n \hat{\alpha}_j y_j \mathbf{x}_i^T \mathbf{x}_j \end{aligned} \quad (29)$$

となり、切片 $\hat{\gamma}$ を求めることができる。

参考文献

[1] 杉山 将. イラストで学ぶ機械学習. 講談社, 2015

[2] サポートベクターマシン (SVM). <http://www.sist.ac.jp/~kanakubo/research/neuro/supportvectormachine.html>

[3] SVM 支持向量机入及数学原理. https://blog.csdn.net/qq_35992440/article/details/80987664