★主に以下の動画にて学習

https://www.youtube.com/watch?v=cNEhKEb9-JU
https://www.youtube.com/watch?v=2IB7vkfGeAA
https://www.youtube.com/watch?v=fUSO6-0o3ds

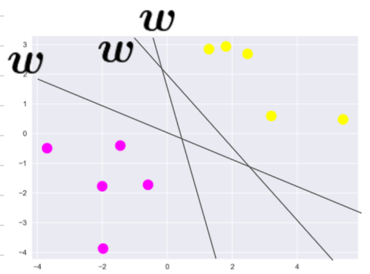
サポートベクターマシン（SVM）

「Support Vector」とは、データを分割する直線に最も近いデータ点の事
サポートベクターを定めてこのような分割線が決まれば、あとはコレより上にあるか下にあるかで、どのクラスに属しているかの予測が出来る。
SVMの肝は、マージン最大化とカーネル法により非線形データを扱えるところ。

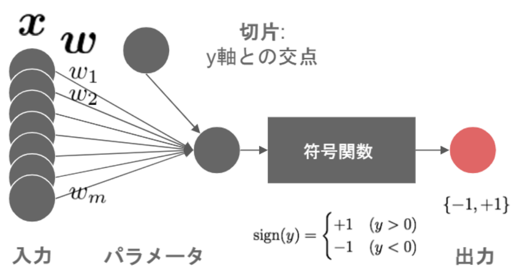
- **SVMのメリット**
  - データの次元が大きくなっても識別精度が良い
  - 最適化すべきパラメータが少ない

- **SVMのデメリット**
  - 学習データが増えると計算量が膨大になる
  - 基本的に2クラス分類に特化している
  - スケーリングが必要
  (SVMでは距離を測定するので、大きい範囲をとる特徴量に引きずられないようにする)

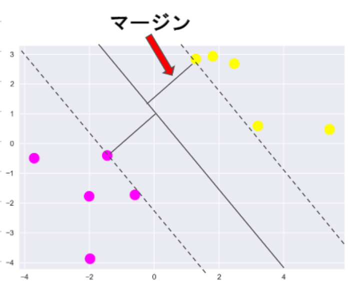
$$y = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b = \sum_{i=1}^{m} w_j x_j + b$$

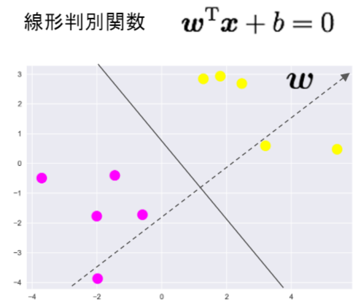


決定境界はいくつも考えられる



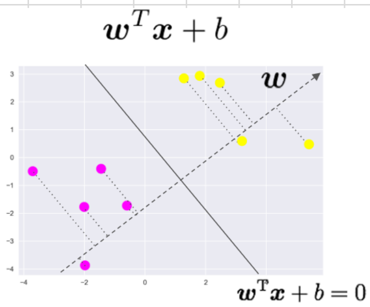
●「**線形判別関数ともっとも近いデータ点との距離**」を「**マージン**」という
●マージンが最大となる線形判別関数を求める

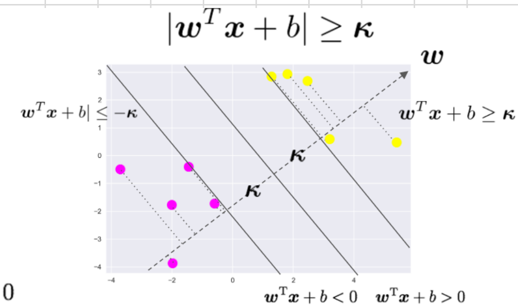




マージンをパラメータに依存する関数として表現！

■目的関数の導出(準備)





各クラスのデータをwの方向κへ射影した点
w軸に座標を変換

線形判別関数のマージンをκとした時に
全てのデータ点でなりたつ条件
(特に路肩上のデータに対しては=κが成り立つ)

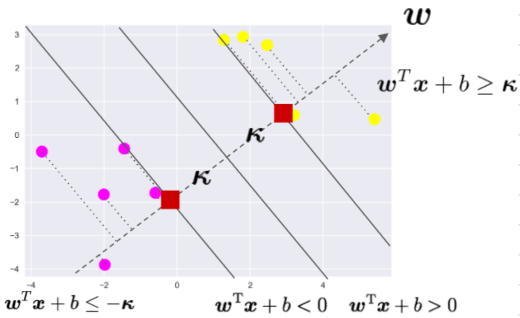
標準座標系に修正！

$$|\boldsymbol{w}^T\boldsymbol{x}+b| \geq \boldsymbol{\kappa}$$

$$\rho(\boldsymbol{w},b) = \min_{\boldsymbol{x}\in C_{t_i=+1}} \frac{\boldsymbol{w}^T\boldsymbol{x}_i}{||\boldsymbol{w}||} - \max_{\boldsymbol{x}\in C_{t_i=-1}} \frac{\boldsymbol{w}^T\boldsymbol{x}_i}{||\boldsymbol{w}||}$$

$$= \frac{1-b}{||\boldsymbol{w}||} - \frac{-1-b}{||\boldsymbol{w}||}$$

$$= \frac{2}{||\boldsymbol{w}||}$$



wが決まればマージンが決まる

https://manabitimes.jp/math/857

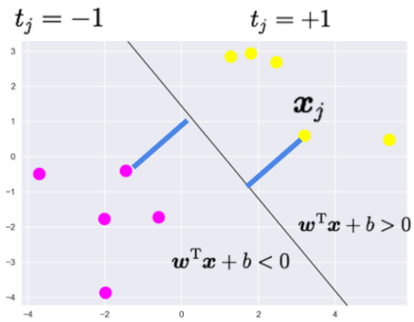
点 $A(x_0, y_0)$ と直線 $l : ax+by+c=0$ の距離 d は,

$$d = \frac{|ax_0+by_0+c|}{\sqrt{a^2+b^2}}$$

各点と決定境界との距離は、点と直線との距離の公式から
例えば：https://mathtrain.jp/tentotyokusen

$$\frac{|\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b|}{||\boldsymbol{w}||} = \frac{t_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)}{||\boldsymbol{w}||}$$

絶対値の定義より



決定境界
$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b=0$

マージンとは決定境界と最も距離の近い点との距離なので、

$$\min_i \frac{t_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)}{||\boldsymbol{w}||}$$

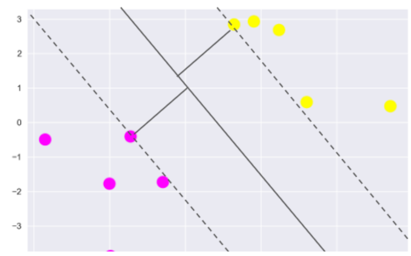
SVMはマージンを最大化することを目標なので目的関数は
、

$$\max_{\boldsymbol{w},b}[\min_i \frac{t_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)}{||\boldsymbol{w}||}]$$

マージンが最大となる直線(パラメータ)を探す

$t_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)=1$ 　 $t_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)=1$

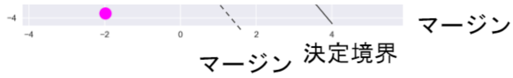
$t_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)>1$ 　 $t_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)>1$



マージン上の点において、$t_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)=1$
が成り立つため、すべての点に対して、
$t_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b) \geq 1$ が成り立ち、目的関数は
以下のようになる。

$$\max_{\boldsymbol{w},b} \frac{1}{||\boldsymbol{w}||}$$

マージンで正規化すると成り立つ



サポートベクターマシンにおける主問題と双対問題

SVMの主問題

・主問題の目的関数と制約条件

$$\min_{\boldsymbol{w},b} \frac{1}{2}||\boldsymbol{w}||^2$$

$$t_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geq 1 \qquad (i = 1, 2, ..., n)$$

上の最適化問題をラグランジュ未定乗数法で解く

ラグランジュの未定乗数法

・制約付き最適化問題を解くための手法

> 制約 $g_i(\boldsymbol{x}) \geq 0 (i = 1, 2, ..., n)$ のもとで、$f(\boldsymbol{x})$ が最小となる $\boldsymbol{x}$ は、
>
> 変数 $\lambda_i \geq 0 (i = 1, 2, ..., n)$ を用いて新たに定義したラグランジュ関数 $L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) - \sum_{i=1}^{n} \lambda_i g(\boldsymbol{x})$ において $\frac{\partial L}{\partial x_j} = 0 \ (j = 1, 2, ..., m)$ を満たす

カルーシュ・クーン・タッカー条件（KKT条件）

KKT条件とは、**「非線形計画において一階導関数が満たすべき最適条件」**を指す。

※**「ラグランジュの未定乗数法」が等式制約のみを扱う**のに対して、**「KKT条件を用いた解法」は不等式制約も扱うことができる。**

KKT条件

・制約付き最適化問題において最適解が満たす条件

制約 $g_i(\boldsymbol{x}) \geq 0 (i = 1, 2, ..., n)$ のもとで、$f(\boldsymbol{x})$ が最小となる $\boldsymbol{x}^*$ は、以下の条件を満たす

$$\frac{\partial L}{\partial x_j}|_{\boldsymbol{x}^*} = 0 \ (j = 1, 2, ..., m)$$

$$g_i(\boldsymbol{x}^*) \leq 0, \ \lambda_i \geq 0, \ \lambda_i g_i(\boldsymbol{x}^*) = 0 \ (i = 1, 2, ..., n)$$

不等式制約でも解けるようになる

（Lはラグランジュ関数、λはラグランジュ乗数である）

ラグランジュの未定乗数法

ラグランジュ関数 $$L(\boldsymbol{w}, b, \boldsymbol{a}) = \frac{1}{2}||\boldsymbol{w}||^2 - \sum_{i=1}^{n} a_i \left(t_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) - 1\right)$$

$a_i \geq 0 (i = 1, 2, ..., n)$ はラグランジュ乗数である。
そして、最小となる $\boldsymbol{w}, b$ は以下を満たす。

$$\frac{\partial L}{\partial \boldsymbol{w}} = \boldsymbol{w} - \sum_{i=1}^{n} a_i t_i \boldsymbol{x}_i = \boldsymbol{0}$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{n} a_i t_i = 0$$

$$\boldsymbol{w} = \sum_{i=1}^{n} a_i t_i \boldsymbol{x}_i$$

$$\sum_{i=1}^{n} a_i t_i = 0$$

不明点：???wがゼロになってしまうが。。。。

双対問題

・ラグランジュ関数から$\boldsymbol{w}, b$を消去

$$\begin{aligned}
\tilde{L} &= \frac{1}{2}||\boldsymbol{w}||^2 - \sum_{i=1}^{n} a_i \left(t_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) - 1\right) \\
&= \frac{1}{2}(\sum_{i=1}^{n} a_i t_i \boldsymbol{x}_i)^{\mathrm{T}}(\sum_{j=1}^{n} a_j t_j \boldsymbol{x}_j) - \sum_{i=1}^{n} a_i t_i (\sum_{j=1}^{n} a_j t_j \boldsymbol{x}_j)^{\mathrm{T}} \boldsymbol{x}_i - b\sum_{i=1}^{n} a_i t_i + \sum_{i=1}^{n} a_i \\
&= \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j t_i t_j \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j
\end{aligned}$$

・双対問題の目的関数と制約条件

$$\max_{\boldsymbol{a}} \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j t_i t_j \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j$$

$$\sum_{i=1}^{n} a_i t_i = 0$$

$$a_i \geq 0 \qquad (i = 1, 2, ..., n)$$

主問題と双対問題

・主問題の最適解と双対問題の最適解は一対一対応

$$\boldsymbol{w} = \sum_{i=1}^{n} a_i t_i \boldsymbol{x}_i$$

$$b = \frac{1}{t_i} - \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i$$

（KKT条件の相補性条件 $a_i\left(t_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) - 1\right) = 0$ より
$a_i > 0$ となる i に対して ）

■予測

予測

・SVMの決定関数

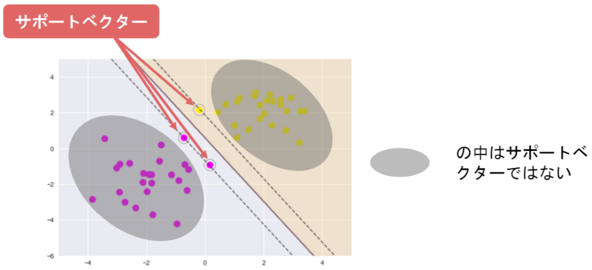
$$y(\boldsymbol{x}) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b = \sum_{i=1}^{n} \underline{a_i} t_i \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} + b$$

KKT条件の相補性条件 $a_i\left(t_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) - 1\right) = 0$ から、

- $t_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b - 1) > 0$ のとき、つまりマージンの外側のデータでは $a_i = 0$ となり予測に影響を与えない。
- $t_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b - 1) = 0$ のとき、つまりマージン上にあるデータでは $a_i > 0$

となり、予測に影響を与える。(サポートベクター)

・分離超平面を構成する学習データは、サポートベクターだけで残りのデータは不要



ハードマージンSVM:線形分離可能な場合

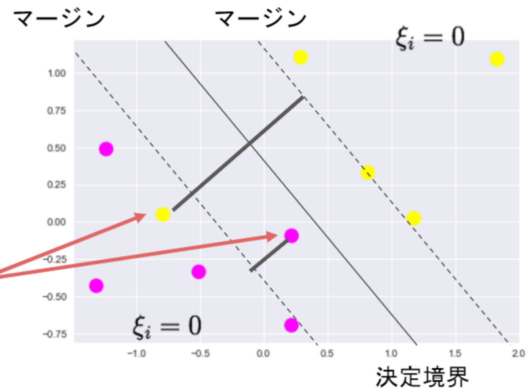
ソフトマージンSVM

- サンプルを線形分離できないとき
- 誤差を許容し、誤差に対してペナルティを与える

ξ:クサイ

マージン内に入るデータや誤分類されたデータに対して誤差を表す変数 ξ_i を導入する

$$\xi_i = 1 - t_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b) > 0$$



ソフトマージンSVMの目的関数と制約条件

ソフトマージンSVMの目的関数と制約条件は以下のようになる

$$\underbrace{\frac{1}{2}||\boldsymbol{w}||^2}_{\text{マージンの大きさ}} + C \underbrace{\sum_{i=1}^{n} \xi_i}_{\text{誤差に対するペナルティ}}$$

トレードオフを制御するパラメータ

双対表現

$$\max_{a} \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j t_i t_j \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j$$

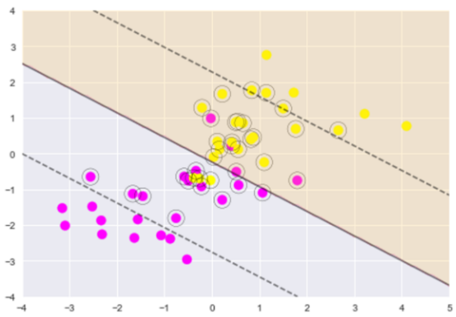
$$\sum_{i=1}^{n} a_i t_i = 0$$

$$0 \le a_i \le C \qquad (i = 1, 2, ..., n)$$

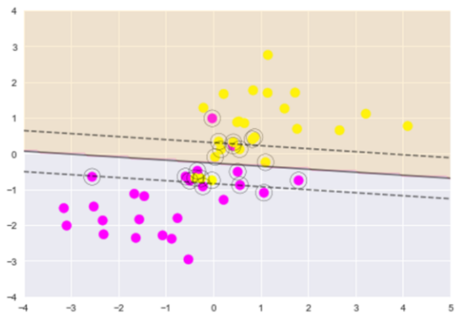
$$t_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b) \ge 1 - \xi_i, \qquad \xi_i \ge 0$$

ソフトマージンSVM

- 線形分離できない場合でも対応
- パラメータCの大小で決定境界が変化



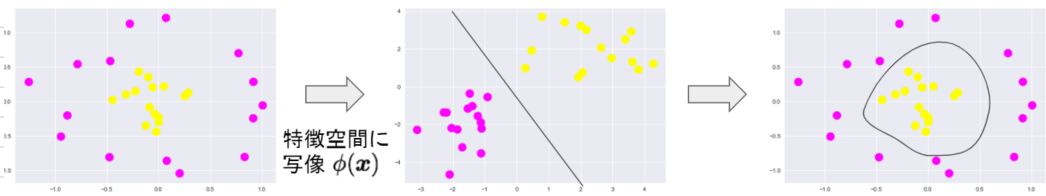
Cが小さい時（誤差をより許容する）



Cが大きい時（誤差をあまり許容しない）

非線形分離

・線形分離できないとき
・特徴空間に写像し、その空間で線形に分離する



線形に分離できない

特徴空間上では
線形に分離可能

もとの空間上では
非線形な分離

カーネルトリック

目的関数は以下のように変わる

$$\max_{\boldsymbol{a}} \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j t_i t_j \underline{\phi(\boldsymbol{x}_i)^{\mathrm{T}}\phi(\boldsymbol{x}_j)}$$

ここのみ変化

カーネルトリック

・カーネル関数 $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i)^{\mathrm{T}}\phi(\boldsymbol{x}_j)$

・高次元ベクトルの内積をスカラー関数で表現

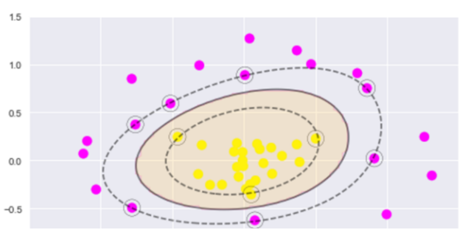
・特徴空間が高次元でも計算コストを抑えられる

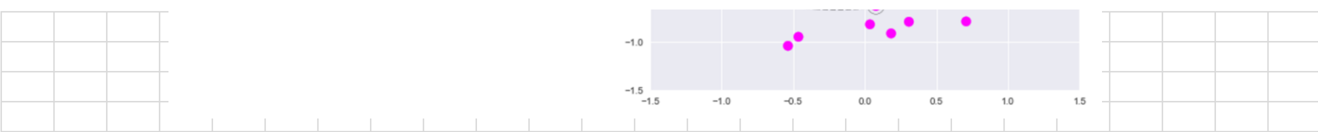
非線形カーネルを用いた分離

・非線形な分離が可能

放射基底関数カーネル（RBFカーネル・ガウシアンカーネル）を用いる

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{||\boldsymbol{x}_i - \boldsymbol{x}_j||^2}{2\sigma^2}\right)$$




−1.0
−1.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5