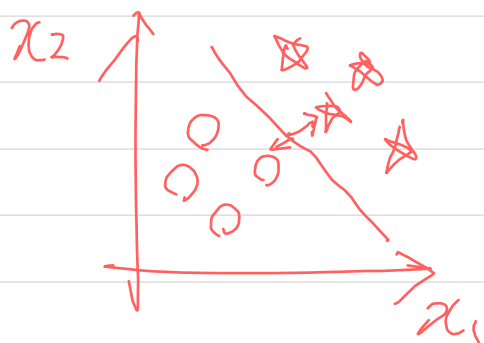


17A

$$x_2 = -\frac{w_1}{w_2} x_1 - \frac{b}{w_2} \rightarrow \begin{matrix} \uparrow \\ \text{y方向に} -w_1 \\ \text{x方向に} -w_2 \end{matrix}$$

Q1

$$f(x_1, x_2) = w_1 x_1 + w_2 x_2 + b$$



$$= w x + b$$

$$\text{where, } x = (x_1, x_2), \quad w = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

$$f(x) = 0, \quad \therefore P(p_1, p_2) \text{ 点 } f(x)$$

における垂線の足を $H(h_1, h_2)$ とすると PH の長さは

$f(x)$ に対する法線ベクトルを n とすると

$$n = (w_1, w_2)$$

$$\overrightarrow{PH} = \alpha \vec{n}$$

$$\Leftrightarrow (h_1 - p_1, h_2 - p_2) = \alpha (w_1, w_2)$$

$$\Leftrightarrow (h_1 - p_1, h_2 - p_2) \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \alpha (w_1, w_2) \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

$$\Leftrightarrow w_1(h_1 - p_1) + w_2(h_2 - p_2) = \alpha (w_1^2 + w_2^2)$$

問題 6 $(w_1^2 + w_2^2)$ を分母と

$$\alpha = \frac{w_1(h_1 - p_1) + w_2(h_2 - p_2)}{(w_1^2 + w_2^2)}$$

よって $\vec{PH}$ の長さは $|\vec{PH}| = \alpha |\vec{u}|$ である)

$$|\vec{PH}| = |\alpha| |\vec{u}|$$

$$= \alpha \times \sqrt{w_1^2 + w_2^2}$$

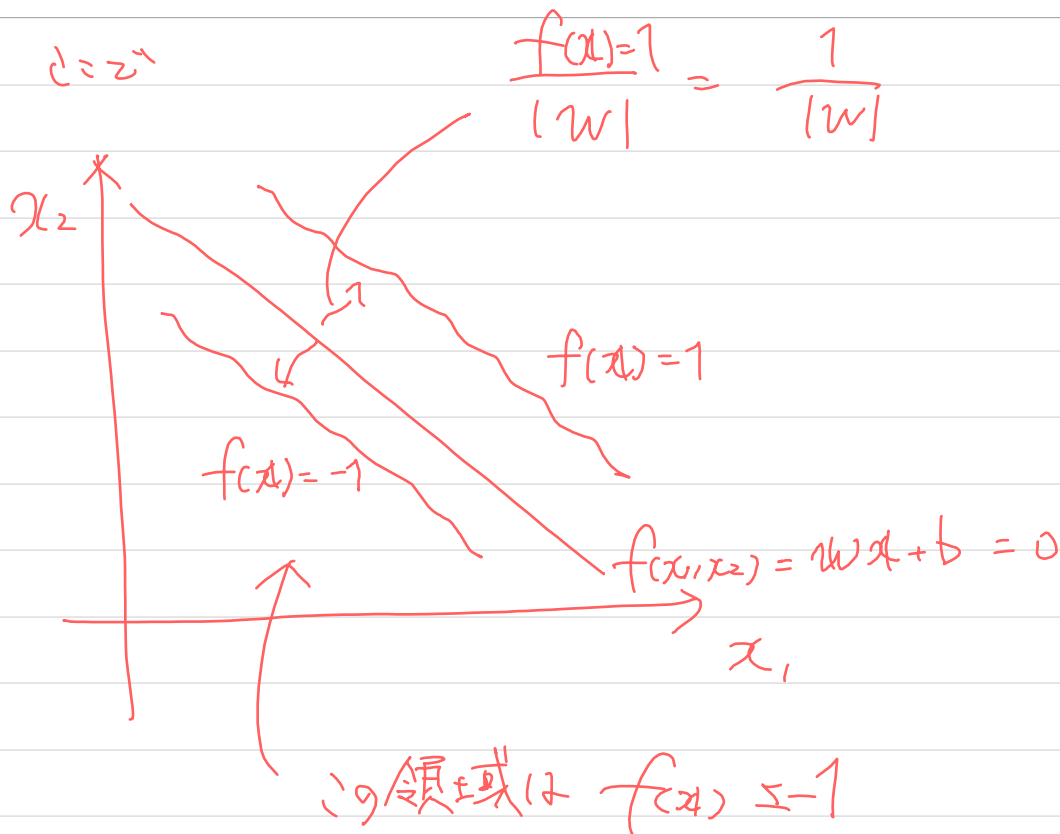
$$= \frac{|w_1(h_1 - p_1) + w_2(h_2 - p_2)|}{\sqrt{w_1^2 + w_2^2}}$$

$H(h_1, h_2)$ (すなわち $f(x_1, x_2) = w_1 x_1 + w_2 x_2 + b = 0$)
上の点である

$$|\vec{PH}| = \frac{|w_1(p_1 - h_1) + w_2(p_2 - h_2)|}{\sqrt{w_1^2 + w_2^2}}$$

$$= \frac{|w_1 p_1 + w_2 p_2 + b|}{\sqrt{w_1^2 + w_2^2}} = \frac{|f(p_1, p_2)|}{\sqrt{w_1^2 + w_2^2}}$$

$$\text{つまり } |\vec{PH}| = \frac{f(p_1, p_2)}{\|w\|}$$



つまり マージン最大化 できるときは $\arg \max_w \left| \frac{2}{\|w\|} \right|$

$\Rightarrow$ 最小化問題 となる $\arg \min_w \frac{\|w\|^2}{2}$

$$\begin{aligned} \text{s.t. } wx_{\text{negative}} + b &\leq -1 \\ wx_{\text{positive}} + b &\geq 1 \end{aligned} \quad \} *$$

$z = z^*$ positive のとき $y_i = 1$
negative のとき $y_i = -1$ となる

* は $y_i (wx_i + b) \geq 1$

パラメータ推定 (最適化)

ラグランジュの未定乗数法によるパラメータ w の推定

$$\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$$

$$L_p(w, b, \lambda) = \frac{\|w^2\|}{2} - \sum_{i=1}^n \lambda_i (y_i (b + wx_i) - 1)$$

つまり) $\arg\min_w L_p \quad \text{s.t.} \quad \lambda_i \geq 0 \quad (i=0, 1, \dots, n)$

$$\frac{\partial L_p}{\partial w} = w - \sum_{i=1}^n \lambda_i y_i x_i = 0 \quad \dots \quad (1)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^n \lambda_i y_i = 0 \quad \dots \dots (2)$$

双対定理より)

$$\arg\max_x C^T x \quad \text{s.t.} \quad Ax = b, x \geq 0$$

$$\arg\min_y b^T y \quad \text{s.t.} \quad A^T x \leq c, y \geq 0$$

$$L_D(\lambda) = \underbrace{\sum_{i=1}^n \lambda_i}_{(2)} - \underbrace{\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j y_i y_j x_i x_j}_{(1)}$$

$$L_p(\lambda) = \underbrace{- \sum_{i=1}^n \lambda_i (y_i (b + wx_i) - 1)}_{(2)} + \underbrace{\frac{1}{2} \|w^2\|}_{(1)}$$

制約

$$0 \leq \lambda_i, \quad \sum_{i=1}^n \lambda_i y_i = 0 \text{ の条件}$$

$$\arg \max_{\lambda} L_D(\lambda), \quad \leftarrow \text{2次最適化問題を解く} \\ \text{ } \lambda \text{ を求める}$$

$$\Rightarrow w \text{ を求める} \quad w = \sum_{i=1}^n \lambda_i y_i x_i^T$$

KKT 条件 (Karush-Kuhn-Tucker condition)

$$f(x) = w^T x + b$$

$$w = \sum_{i=1}^n \lambda_i y_i x_i \quad \# \text{)}$$

$$f(\hat{x}) = \sum_{i=1}^n \sum_{j=1}^n \hat{\lambda}_i y_i x_i^T x_j + b$$

$$b = y_i - \sum_{i=1}^n \sum_{j=1}^n \hat{\lambda}_i y_i x_i^T x_j$$

$$\theta = \frac{1}{N_S} \sum_{n \in S} \left(y_n - \sum_{m \in S} \hat{\lambda}_m y_m x_m^T x_n \right)$$

$S = \{1, 2, \dots, N\}$ の集合

$$\operatorname{argmax}_{\lambda} \quad \sum_{i=1}^n \lambda_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j y_i y_j x_i x_j = L(w, b, \lambda)$$

$$\text{st} \quad \sum_{i=1}^n \lambda_i y_i = 0 \quad \lambda_i \geq 0$$

凸2次計画

$$\lambda_i^{(k+1)} = \lambda_i^{(k)} + \frac{\partial L(w, b, \lambda)}{\partial \lambda}$$

$$\lambda_i^{(k+1)} = \lambda_i^{(k)} + \alpha \left\{ 1 - \sum_{j=1}^n \lambda_j y_j x_j \right\}$$

SGDとかの反復解法で

λ を求める

$$\text{条件} \quad \sum_{i=1}^n \lambda_i y_i = 0, \quad \lambda_i \geq 0 \quad (*) \quad y_i \neq 0$$

とすると、

$$\text{また} \quad w = \sum_{i=1}^n \lambda_i y_i x_i$$

もし w が求まれば

おしまい

カーネル法の特長

$k(x_1, x_2)$ を 条件付きで適宜に決めると

↳ e.g. ガウスカーネル $= \exp\left(-\frac{|x-y|^2}{2\sigma^2}\right)$

は 無限個の特徴量 $\phi = \{\phi_1(x), \phi_2(x_2), \dots\}$

を使っていると同じ

SVMを定式化

$$f(x) = \sum_{n=1}^N \alpha_n \underline{k(x, x_n)}$$

$$\hookrightarrow k(x, y) = \sum_{m=1}^M \phi_m(x) \phi_m(y)$$

$\underbrace{O(n \times m)}_{\text{ベクトル}} \text{ の } \underbrace{\text{dot-product}}_{\text{特徴量}}$

$k(x, y)$ を 適宜に置き換え ること

↳ $\phi_1(x), \dots, \phi_n(x), \dots$ 無限次元の特徴量がい
まろーソに展開できると

$\Rightarrow O(n)$ の dot-product に

$\delta(x, y)$ は 2点の距離が 近いか遠い値

$\Rightarrow \delta_n$ は ラベル $+1$ or -1 に符号が 一致する.

0に近いと 近距離の $\delta=1$ と同じ

いざこれを Overfitting してしまう. なので σ を 適度に

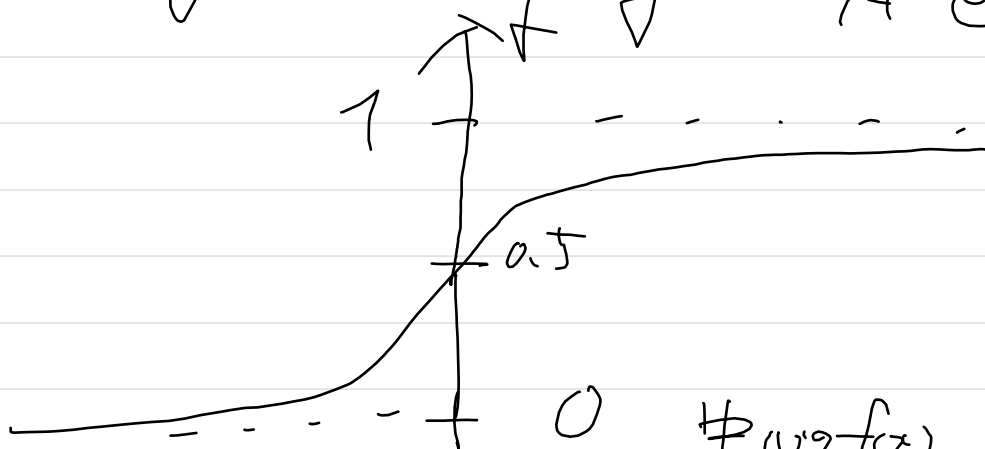
調整 する必要がある, いわゆる Hyper-Param-Tuning.

前ノシテハ 赤ノ書ニテハ 黒ニテハ

softmax
$$y_i = \frac{e^{x_i}}{\sum_{j=1}^N e^{x_j}}$$

$0 \leq y_i \leq 1$ 満足す

いかなる sigmoid の $y = \frac{1}{1 + e^{-x}}$



一般化 logistic function

$$f(x) = \frac{L}{1 + \exp(-\lambda(x - x_0))}$$

max 値

$L = 1$
 $x_0 = 0$
 $\lambda = 1$
 一般化 logistic function

Sigmoidに $-\infty, \infty$

$\downarrow$
 $0, 1$ に map する

cumulative distribution function

累積分布関数の性質と同様.

では softmax は?

Sigmoid のような

softmax は ベクトル への

正規化 = ベクトルを返す

$\downarrow$

出力で 7 個の確率を返す $\rightarrow$ 分類問題に使う.

softmaxの Overflow or Underflow



$$\text{softmax}(x)_k = y_k =$$

$$\frac{\exp(x_k)}{\sum_{i=1}^C \exp(x_i)}$$

$(1 \leq k \leq C \text{ (1以上の整数)})$

$$\exp(x_k) = e^{x_k}$$

もし x_k が大きすぎると
 e^{x_k} は すぐ ∞ になる

同様に α をかけると $y_k = \frac{\alpha \exp(x_k)}{\alpha \sum_{i=1}^C \exp(x_i)}$

$$= \frac{\exp(x_k + \log \alpha)}{\sum_{i=1}^C \exp(x_i + \log \alpha)}$$

$\therefore \alpha = \max(x)$ とすれば正規化でき
 なる