习题 3-1

决策平面为

$$f(x,w) = w^T x + b = 0$$

取平面上任意两点 x_1, x_2, 构成一向量 $x_1 - x_2$, 则

$$\begin{aligned} w \cdot (x_1 - x_2) =& w^T(x_1 - x_2) \\ =& w^T x_1 - w^T x_2 \\ =& (w^T x_1 + b) - (w^T x_2 + b) \\ =& 0 \end{aligned}$$

故 w 与决策平面正交

习题 3-2

$$\begin{aligned} d =& \frac{|w^T x + b|}{\sqrt{\sum_i w_i^2}} \\ =& \frac{|f(x,w)|}{\|w\|} \end{aligned}$$

习题 3-3

令 $X = [x_1, x_2, \cdots, x_n, 1]^T$, $W^\star = [w_1^\star, w_2^\star, \cdots, b^\star]^T$, 因为 $yf(x, w^\star) > 0, \forall x$ 我们有

$$Y^T(X^T W^\star) > 0$$

令 $Y^T(X^T W^\star) = p, p > 0$, 则

$$\begin{aligned} Y^T X^T w^\star =& p \\ w^\star =& p(Y^T X^T)^{-1} \end{aligned}$$

故 $w^\star$ 为 X 的线性组合

习题 3-4

两类线性可分:

$$\begin{aligned} yf((\rho x_1+(1-\rho)x_2),w) =& y\big(w^T(\rho x_1+(1-\rho)x_2)+b\big) \\ =& y\big(w^T\rho x_1+w^T(1-\rho)x_2+b\big) \\ =& y\big(w^T\rho x_1+\rho b\big)+y\big(w^T(1-\rho)x_2+(1-\rho)b\big) \\ =& \rho y\big(w^Tx_1+b\big)+(1-\rho)y\big(w^Tx_2+b\big) \\ >& 0 \end{aligned}$$

即, $yf((\rho x_1+(1-\rho)x_2),w)>0$

多类线性可分:

$$\begin{aligned} f_c(\rho x_1+(1-\rho)x_2) =& w_c^T\big(\rho x_1+(1-\rho)x_2\big)+b_c \\ =& \rho\big(w_c^Tx_1+b_c\big)+(1-\rho)\big(w_c^Tx_2+b_c\big) \\ >& \rho\big(w_{\tilde{c}}^Tx_1+b_{\tilde{c}}\big)+(1-\rho)\big(w_{\tilde{c}}^Tx_2+b_{\tilde{c}}\big) \\ =& w_{\tilde{c}}^T\big(\rho x_1+(1-\rho)x_2\big)+b_{\tilde{c}} \\ =& f_{\tilde{c}}((\rho x_1+(1-\rho)x_2),w_{\tilde{c}}) \end{aligned}$$

即, $f_{\tilde{c}}(\rho x_1+(1-\rho)x_2)>f_{\tilde{c}}((\rho x_1+(1-\rho)x_2),w_{\tilde{c}})$

习题 3-5

(1)

取任意类 c, 任意类 $\tilde{c}(\forall\tilde{c}\neq c)$, x_c 为一 c 类样本
因为类 c 对其它类线性可分, 故 $f(x_c,w_c^\star)>f(x_c,w_{\tilde{c}}^\star)$
显然, 对所有 C 个类, 都存在一个权重向量 $w_c^\star, 1\leq c\leq C$, 使得 $f(x_c,w_c^\star)>f(x_c,w_{\tilde{c}}^\star),\forall\tilde{c}\neq c$ 成立
故训练集线性可分

(2)

如图 1所示, 叉, 三角, 圆为不同的三类, 其中两两线性可分, 但该数据集并不是线性可分的

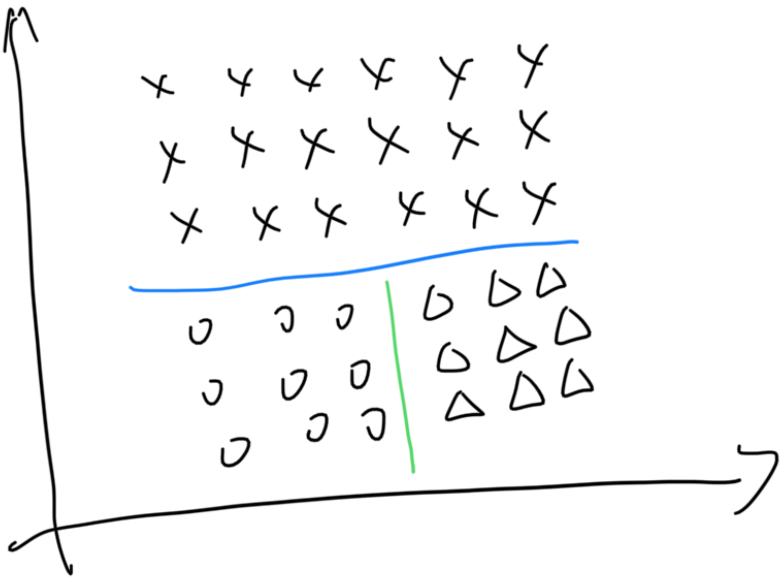

图 1: 每两个类线性可分

习题 3-6

可以

$$
\begin{aligned}
\left(y-\hat{y}\right)^2 =& \left(y-\sigma(f(x,w))\right)^2 \\
=& y^2 - 2y\sigma(f(x,w)) + \sigma^2(f(x,w)) \\
=& 1 - 2y\sigma(f(x,w)) + \left(y\sigma(f(x,w))\right)^2 \\
=& \left(1 - y\sigma(f(x,w))\right)^2
\end{aligned}
$$

因为 $\sigma(f(x,w))$ 的取值范围为 $(0,1)$, 故 $y\sigma(f(x,w))$ 的取值范围为 $(-1,1)$.
损失函数 $\mathcal{L} = \left(1 - y\sigma(f(x,w))\right)^2$ 在 $(-1,1)$ 区间内单调递减, 所以可以使用.

习题 3-9

根据学习算法, 我们有

$$
w_k = \sum_{k=1}^{K} \left(\phi(x^{(k)}, y^{(k)}) - \phi(x^{(k)}, \hat{y}^{(k)})\right)
$$

计算 $\|w_k\|^2$ 的上界

$$
\begin{aligned}
\|w_k\|^2 =& \|w_{k-1} + (\phi(x^{(k)}, y^{(k)}) - \phi(x^{(k)}, \hat{y}^{(k)}))\|^2 \\
=& \|w_{k-1}\|^2 + \|\phi(x^{(k)}, y^{(k)}) - \phi(x^{(k)}, \hat{y}^{(k)})\|^2 + 2 < w_{k-1}, \phi(x^{(k)}, y^{(k)}) - \phi(x^{(k)}, \hat{y}^{(k)}) > \\
\leq& \|w_{k-1}\|^2 + \|\phi(x^{(k)}, y^{(k)}) - \phi(x^{(k)}, z)\|^2 \\
=& \|w_{k-1}\|^2 + R^2 \\
\leq& KR^2
\end{aligned}
$$

计算 $\|w_k\|^2$ 的下界

$$
\begin{aligned}
\|w_k\|^2 =& 1 \cdot \|w_{k-1}\|^2 \\
=& \|w^\star\| \cdot \|w_{k-1}\|^2 \\
\geq& \|w^{\star T} w_k\|^2 \\
=& \|w^{\star T} \sum_{k=1}^{K} (\phi(x^{(k)}, y^{(k)}) - \phi(x^{(k)}, \hat{y}^{(k)}))\|^2 \\
=& \| \sum_{k=1}^{K} < w^{\star T}, \phi(x^{(k)}, y^{(k)}) - \phi(x^{(k)}, \hat{y}^{(k)}) > \|^2 \\
\geq& K^2 \gamma^2
\end{aligned}
$$

最终,

$$
\begin{aligned}
KR^2 \geq& K^2 \gamma^2 \\
K \leq& \frac{R^2}{\gamma^2}
\end{aligned}
$$