

Ch3 线性模型

习题3-1

证明在两类线性分类中，权重向量与决策平面正交。

答：

法1:

$$\begin{aligned}\text{线性判别函数为 } f(x; w) &= w_1 x_1 + w_2 x_2 + \dots + w_D x_D + b \\ &= w^T x + b\end{aligned}$$

$$x = [x_1, \dots, x_D]^T$$

$$\text{权重向量为 } w = [w_1, \dots, w_D]^T$$

$$\text{决策平面为 } f(x; w) = 0$$

法1: 取 $x_1, x_2 \in f(x; w) = 0$ ，则

$$w^T x_1 + b = 0$$

$$w^T x_2 + b = 0$$

$$\longrightarrow w^T (x_1 - x_2) = 0$$

因为 $(x_1 - x_2) \parallel$ 决策平面，所以 w 垂直于 $(x_1 - x_2)$ ，权重向量与决策平面正交。

法2:

平面 $Ax + By + Cz = D$ 的法向量为 $[A \ B \ C]$ ，同理平面 $w^T x + b$ 的法向量为 w

习题3-2

在线性空间中，证明一个点 x 到平面 $f(x; w) = w^T x + b = 0$ 的距离为 $|f(x; w)| / \|w\|$ 。

答：

法1:

设 x 在平面 $f(x; w)$ 上的投影为点 x' ，到平面的距离为 $\|x - x'\|$

因为 $x - x'$ 与 w 平行， $\exists k \in R$ 使得 $x - x' = kw$ ， $\|x - x'\| = |k| \|w\|$

$$f(x; w) - f(x'; w) = w^T (x - x')$$

$$\text{又 } f(x'; w) = w^T x' + b = 0,$$

$$\text{所以 } f(x; w) - f(x'; w) = f(x; w) = w^T (x - x')$$

将 $x - x' = kw$ 带入上式，得 $f(x; w) = w^T kw$

$$\text{即 } f(x; w) = k \|w\|^2, \text{ 推出 } |k| = \frac{|f(x; w)|}{\|w\|^2} \text{ 带入}$$

$$\text{到平面到距离 } \|x - x'\| = |k| \|w\| = |f(x; w)| / \|w\|$$

法2:

设 x 在平面 $f(x; w)$ 上的投影为点 x' , 向量 $\overrightarrow{xx'}$ 为以点 x 为起点, x' 为终点的有向向量

w 是平面的一个法向量

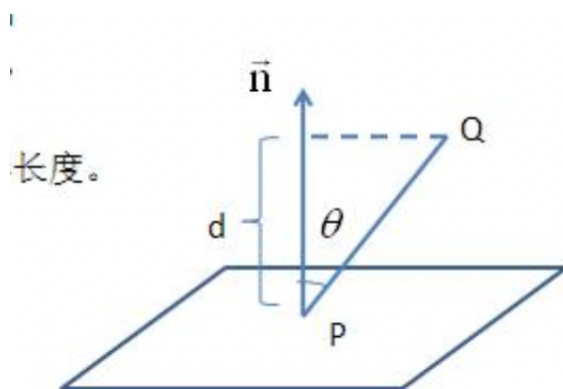
$$d = |\overrightarrow{xx'}| \cos \theta = \frac{w^T(x' - x)}{\|w\|}$$

又 $f(x'; w) = w^T x' + b = 0$, 推出 $w^T x' = -b$



$$d = |\overrightarrow{xx'}| \cos \theta = \frac{w^T(x' - x)}{\|w\|} = \frac{-b - w^T x}{\|w\|}$$

$$d = \frac{|-b - w^T x|}{\|w\|} = \frac{|f(x; w)|}{\|w\|}$$



习题3-3

在线性分类中, 决策区域是凸的。即若点 x_1 和 x_2 被分为类别 c , 则点 $\rho x_1 + (1 - \rho)x_2$ 也会被分为类别 c , 其中 $\rho \in (0, 1)$ 。

答:

1. 二分类

设判别函数为 $f(x; w) = w^T x + b$, $f(x_1; w) > 0$, $f(x_2; w) > 0$, 且 $x_1 > x_2$

$$\begin{aligned} f(\rho x_1 + (1 - \rho)x_2; w) &= w^T(\rho x_1 + (1 - \rho)x_2) + b \\ &= \rho w^T(x_1 - x_2) + w^T x_2 + b \end{aligned}$$

$$\text{又 } f(x_1; w) - f(x_2; w) = w^T(x_1 - x_2) + b > 0$$

所以 $\rho w^T(x_1 - x_2) + w^T x_2 + b > 0$, 点 $\rho x_1 + (1 - \rho)x_2$ 也会被分为类别 c

2. 多分类

由多分类"argmax"方式, 若点 x_1 和 x_2 被分为类别 c , $f_c(x_1; w_c) > f_{\bar{c}}(x_1; w_{\bar{c}})$, $f_c(x_2; w_c) > f_{\bar{c}}(x_2; w_{\bar{c}})$

则

$$\rho f_c(x_1; w_c) > \rho f_{\bar{c}}(x_1; w_{\bar{c}}), (1 - \rho) f_c(x_2; w_c) > (1 - \rho) f_{\bar{c}}(x_2; w_{\bar{c}})$$

$$\rho(w_c^T x_1 + b_c) - \rho(w_{\bar{c}}^T x_1 + b_{\bar{c}}) > 0, (1 - \rho)(w_c^T x_2 + b_c) - (1 - \rho)(w_{\bar{c}}^T x_2 + b_{\bar{c}}) > 0$$

$$\rho(w_c^T x_1 + b_c) - \rho(w_{\bar{c}}^T x_1 + b_{\bar{c}}) + (1 - \rho)(w_c^T x_2 + b_c) - (1 - \rho)(w_{\bar{c}}^T x_2 + b_{\bar{c}}) > 0$$

$$w_c^T(\rho x_1 + (1 - \rho)x_2) + \rho b_c + (1 - \rho)b_c - w_{\bar{c}}^T(\rho x_1 + (1 - \rho)x_2) - \rho b_{\bar{c}} - (1 - \rho)b_{\bar{c}} > 0$$

$$w_c^T(\rho x_1 + (1 - \rho)x_2) + b_c - w_{\bar{c}}^T(\rho x_1 + (1 - \rho)x_2) - b_{\bar{c}} > 0$$

$$\text{所以 } f_c(\rho x_1 + (1 - \rho)x_2; w_c) > \rho f_{\bar{c}}(\rho x_1 + (1 - \rho)x_2; w_{\bar{c}})$$

点 $\rho x_1 + (1 - \rho)x_2$ 也会被分为类别 c

习题3-4

给定一个多分类的数据集，证明：1) 如果数据集中每个类的样本都和除该类之外的样本是线性可分的，则该数据集一定是线性可分的；2) 如果数据集中每两个类的样本是线性可分的，则该数据集不一定是线性可分的。

答：

1) 定义3.2-多类线性可分：对于训练集 D ，若存在 C 个权重向量 w_1^*, \dots, w_C^* ，使得第 $c(1 \leq c \leq C)$ 类的所有样本都满足 $f_c(x; w_c^*) > f_{\bar{c}}(x; w_{\bar{c}}^*), \forall \bar{c} \neq c$ ，那么训练集 D 是线性可分的。 $f_c(x; w_c) = w_c^T x + b_c$

如题意，假设第 $c(1 \leq c \leq C)$ 类的样本都和除该类之外的样本是线性可分的，则

$$\exists w_c^* \text{ 使得 } f_c(x; w_c^*) > f_{\bar{c}}(x; w_{\bar{c}}^*), \forall \bar{c} \neq c$$

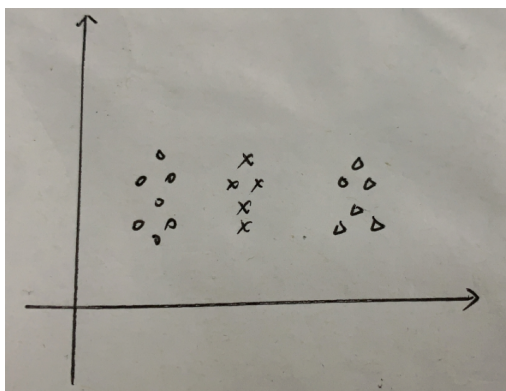
对每一类都成立，即

$$\text{存在 } C \text{ 个权重向量 } w_1^*, \dots, w_C^*, \text{ 使得第 } c(1 \leq c \leq C) \text{ 类的所有样本都满足 } f_c(x; w_c^*) > f_{\bar{c}}(x; w_{\bar{c}}^*), \forall \bar{c} \neq c,$$

$$\sum_{c=1}^C (f_c(x; w_c^*) - f_{\bar{c}}(x; w_{\bar{c}}^*)) > 0$$

则训练集 D 是线性可分的

2)



习题3-5

在Logistic回归中，是否可以用 $\hat{y} = \sigma(w^T x)$ 去逼近正确的标签 y ，并用平方损失 $(y - \hat{y})^2$ 最小化来优化参数 w ?

答：

不能。

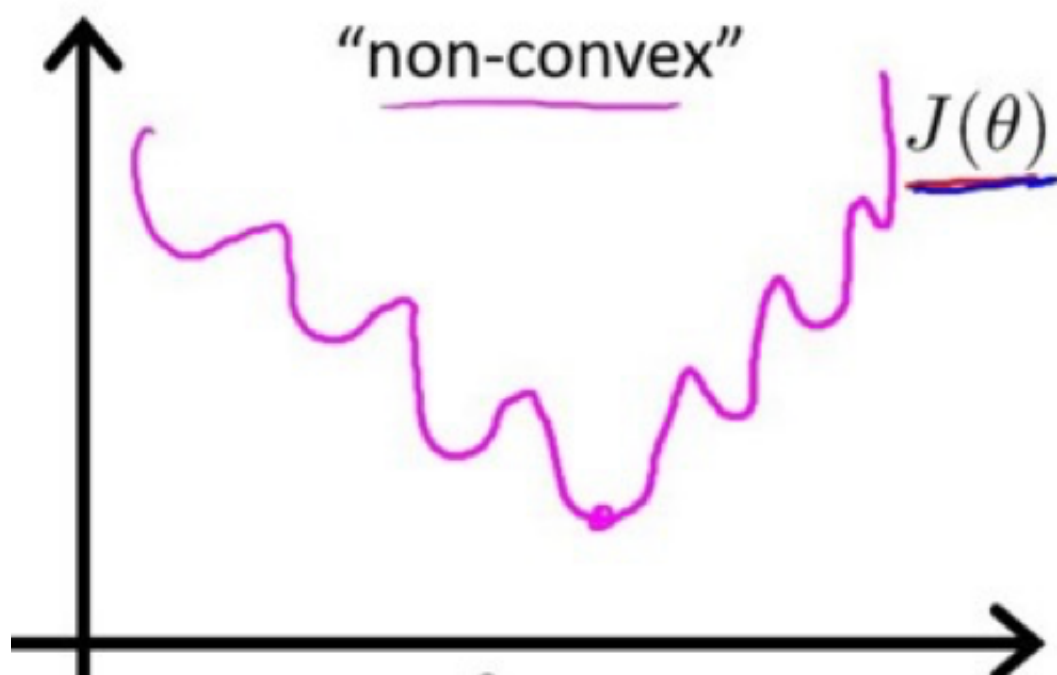
1. Logistic回归采用Logistic函数作为激活函数。

用平方损失 $(y - \hat{y})^2$ 作为损失函数时,

$$L = \frac{1}{2}(\hat{y} - y)^2$$

2.Logistic回归是用于二分类的线性模型, $\hat{y} = \sigma(w^T x)$ 是连续函数, 不适用于分类问题, 应该使用非线性函数将线性函数的值压缩到(0,1)之间。平方损失函数不适用于分类问题, 因为在分类问题中每个标签的距离没有意义。

3.logistic函数的激活函数是sigmoid函数, 是非线性的, 导致损失函数的曲线变得很复杂, 并不是凸函数, 不利于优化, 很容易陷入局部最优解的情况。这时候损失函数曲线如下图所示:



习题3-6

在 Softmax 回归的风险函数 (公式 (3.39)) 中, 如果加上正则化项会有什么影响?

答:

公式 3.39

$$R(w) = -\frac{1}{N} \sum_{n=1}^N (y^{(n)})^T \log \hat{y}^{(n)}$$

$\hat{y}^{(n)} = \text{softmax}(W^T x^{(n)})$ 为样本 $x^{(n)}$ 在每个类别的后验概率

Softmax回归中使用的C个权重向量是冗余的, 即对所有的 权重向量都减去一个同样的向量 v , 不改变其输出结果. 因此, Softmax 回归往往需要使用正则化来约束其参数. 此外, 我们还可以利用这个特性来避免计算Softmax函数时在数值计算上溢出问题。

不加入正则化项限制权重向量的大小, 可能造成权重向量过大, 产生上溢。

补充: 正则化就是对最小化经验误差函数上加约束。约束有引导作用, 在优化误差函数的时候倾向于选择满足约束的梯度减少的方向, 使最终的解倾向于符合先验知识。正则化的作用是选择经验风险函数与模型复杂度同时较小的模型。能够减少噪声的影响, 降低过拟合

习题3-9

若数据集线性可分，证明支持向量机中将两类样本正确分开的最大间隔分割超平面存在且唯一。

答：

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (1)$$

$$s.t. y_i(wx_i + b) - 1 \geq 0, i = 1, 2, 3, \dots, N \quad (2)$$

$$D = \{x_i, y_i\}_{i=1}^N \quad y_i \in \{+1, -1\}$$

1. 首先证明超平面的存在性

由于数据集线性可分，所以公式(1),(2)一定存在可行解，由于目标函数有下界，所以最优化问题必有最优解，记做 (w, b) ;

由于数据集中有正类点和负类点， w 不能为0，因为 w 为0时，意味着不存在超平面能将正类点和负类点分开，所以 w 等于0不是可行解，因而最优解 (w, b) 必定满足 w^* 不能等于0，所以超平面存在

2. 证明超平面存在的唯一性

证明超平面的唯一性即证明 (w, b) 唯一性

假设存在最优解 (w_1^*, b_1^*) 和 (w_2^*, b_2^*) ，则

$$\|w_1^*\| = \|w_2^*\| \quad (3)$$

根据式(2)

$$y_i(w_1^*x_i + b_1^*) - 1 \geq 0 \quad (4)$$

$$y_i(w_2^*x_i + b_2^*) - 1 \geq 0 \quad (5)$$

(4)+(5)得

$$y_i\left(\frac{w_1^* + w_2^*}{2}x_i + \frac{b_1^* + b_2^*}{2}\right) - 1 \geq 0 \quad (6)$$

令 $w = \frac{w_1^* + w_2^*}{2}$, $b = \frac{b_1^* + b_2^*}{2}$ ，可知 w, b 也是最优解，又向量相加的模和向量的模相加有如下关系

$$\|a + b\| \leq \|a\| + \|b\| \quad (7)$$

根据式(3)、(6)、(7)

$$\left\|\frac{w_1^*}{2} + \frac{w_2^*}{2}\right\| \leq \left\|\frac{w_1^*}{2}\right\| + \left\|\frac{w_2^*}{2}\right\|$$

$$\|w\| = \left\|\frac{w_1^*}{2} + \frac{w_2^*}{2}\right\|$$

$$\|w_1^*\| = \|w_2^*\| = \|w\|$$

$$\rightarrow \|w\| = \left\|\frac{w_1^*}{2} + \frac{w_2^*}{2}\right\| \leq \left\|\frac{w_1^*}{2}\right\| + \left\|\frac{w_2^*}{2}\right\| = \frac{\|w_1^*\|}{2} + \frac{\|w_2^*\|}{2} = \|w\| \quad (8)$$

$$\text{由(8), } \left\|\frac{w_1^*}{2} + \frac{w_2^*}{2}\right\| = \left\|\frac{w_1^*}{2}\right\| + \left\|\frac{w_2^*}{2}\right\|$$

$$\text{所以 } w_1^* // w_2^* \iff w_1^* = \lambda w_2^*$$

因为式(3), $\lambda = 1$ or $\lambda = -1$, 又 $\lambda = -1$ 时, $w = 0$, 意味着不存在超平面能将正类点和负类点分开, 所以 $\lambda = 1$, 即 $w_1^* = w_2^*$, w 是唯一的。

下证 b 是唯一的。

最优解变为 (w^*, b_1^*) 和 (w^*, b_2^*)

取超平面 (w^*, b_1^*) 上 2 点 $(x_1, 1), (x_2, -1)$

$$w^* x_1 + b_1^* - 1 = 0 \quad (9)$$

$$-w^* x_2 - b_1^* - 1 = 0 \quad (10)$$

取超平面 (w^*, b_2^*) 上 2 点 $(x_3, 1), (x_4, -1)$

$$w^* x_3 + b_2^* - 1 = 0 \quad (11)$$

$$-w^* x_4 - b_2^* - 1 = 0 \quad (12)$$

由(9)(10)(11)(12)得

$$b_1^* = -\frac{1}{2}w^*(x_1 + x_2) \quad (13)$$

$$b_2^* = -\frac{1}{2}w^*(x_3 + x_4) \quad (14)$$

(13)-(14)得

$$b_1^* - b_2^* = -\frac{1}{2}(w^*(x_1 - x_3) + w^*(x_2 - x_4))$$

将 $(x_1, 1)$ 代入超平面 (w^*, b_2^*) , $(x_3, 1)$ 代入超平面 (w^*, b_1^*)

$$w^* x_1 + b_2^* - 1 \geq 0 \quad (15)$$

$$w^* x_3 + b_1^* - 1 \geq 0 \quad (16)$$

由(9)(11)(15)(16)得

$$w^* x_1 + b_2^* \geq 1 = w^* x_3 + b_2^*$$

$$w^* x_3 + b_1^* \geq 1 = w^* x_1 + b_1^*$$

$$\Rightarrow w^*(x_1 - x_3) \geq 0$$

$$\Rightarrow w^*(x_1 - x_3) \leq 0$$

所以 $w^*(x_1 - x_3) = 0$

将 $(x_2, -1)$ 代入超平面 (w^*, b_2^*) , $(x_4, -1)$ 代入超平面 (w^*, b_1^*)

$$-w^* x_2 - b_2^* - 1 \geq 0 \quad (17)$$

$$-w^* x_4 - b_1^* - 1 \geq 0 \quad (18)$$

由(10)(12)(17)(18)得

$$-w^* x_2 - b_2^* \geq 1 = -w^* x_4 - b_2^*$$

$$-w^* x_4 + b_1^* \geq 1 = -w^* x_2 - b_1^*$$

$$\Rightarrow w^*(x_2 - x_4) \leq 0$$

$$\Rightarrow w^*(x_2 - x_4) \geq 0$$

$$\text{所以 } w^*(x_2 - x_4) = 0$$

$$\text{所以 } b_1^* - b_2^* = -\frac{1}{2}(w^*(x_1 - x_3) + w^*(x_2 - x_4)) = 0, \text{ 即 } b_1^* = b_2^*$$

综上，最优解 (w_1^*, b_1^*) 和 (w_2^*, b_2^*) 相同，只存在一个超平面