

统计机器学习

Statistical Machine learning

Linear Classifier

习题答案

题目：使用感知机算法，求决策函数，权重初始化为零向量。

数据：正类 4 个样本 $\{(1, 1), (1, 2), (2, 2), (1, 0)\}$ ；

负类 4 个样本 $\{(4, 5), (5, 6), (6, 7), (6, 6)\}$ ；

解:

将数据表示成增广向量的形式

$$\{x_1(1, 1, 1), x_2(1, 2, 1), x_3(2, 2, 1), x_4(1, 0, 1), \\ x_5(4, 5, 1), x_6(5, 6, 1), x_7(6, 7, 1), x_8(6, 6, 1)\} \\ w_0 = (0, 0, 0)$$

第一轮迭代：因为 $w_0 x_1^T = 0 \not> 0$ ，所以 $w_0 = w_0 + x_1 = (1, 1, 1)$

因为 $w_0 x_2^T > 0$ ，所以权向量不变。

因为 $w_0 x_3^T > 0$ ，所以权向量不变。

因为 $w_0 x_4^T > 0$ ，所以权向量不变。

因为 $w_0 x_5^T > 0$ ，所以 $w_0 = w_0 - x_5 = (-3, -4, 0)$

因为 $w_0 x_6^T \leq 0$ ，所以权向量不变。

因为 $w_0 x_7^T \leq 0$ ，所以权向量不变。

因为 $w_0 x_8^T \leq 0$ ，所以权向量不变。

$$w_1 = (-3, -4, 0)$$

第二轮迭代：因为 $w_1 x_1^T < 0$ ，所以 $w_1 = w_1 + x_1 = (-2, -3, 1)$

因为 $w_1 x_2^T < 0$ ，所以 $w_1 = w_1 + x_2 = (-1, -1, 2)$

因为 $w_1 x_3^T < 0$ ，所以 $w_1 = w_1 + x_3 = (1, 1, 3)$

因为 $w_1 x_4^T > 0$ ，所以权向量不变。

因为 $w_1 x_5^T > 0$ ，所以 $w_1 = w_1 - x_5 = (-3, -4, 2)$

因为 $w_1 x_6^T \leq 0$ ，所以权向量不变。

因为 $w_1 x_7^T \leq 0$ ，所以权向量不变。

因为 $w_1 x_8^T \leq 0$ ，所以权向量不变。

$$w_2 = (-3, -4, 2)$$

第三轮迭代：因为 $w_2 x_1^T < 0$ ，所以 $w_2 = w_2 + x_1 = (-2, -3, 3)$

因为 $w_2 x_2^T < 0$ ，所以 $w_2 = w_2 + x_2 = (-1, -1, 4)$

因为 $w_2 x_3^T \leq 0$ ，所以 $w_2 = w_2 + x_3 = (1, 1, 5)$ 。

因为 $w_2 x_4^T > 0$ ，所以权向量不变。

因为 $w_2 x_5^T > 0$ ，所以 $w_2 = w_2 - x_5 = (-3, -4, 4)$ 。

因为 $w_2 x_6^T \leq 0$ ，所以权向量不变。

因为 $w_2 x_7^T \leq 0$ ，所以权向量不变。

因为 $w_2 x_8^T \leq 0$ ，所以权向量不变。

$$w_3 = (-3, -4, 4)$$

第四轮迭代：因为 $w_2 x_1^T < 0$ ，所以 $w_3 = w_3 + x_1 = (-2, -3, 5)$

因为 $w_2 x_2^T < 0$ ，所以 $w_3 = w_3 + x_2 = (-1, -1, 6)$

因为 $w_2 x_3^T > 0$ ，所以权向量不变。

因为 $w_2 x_4^T > 0$ ，所以权向量不变。

因为 $w_2 x_5^T \leq 0$ ，所以权向量不变。

因为 $w_2 x_6^T \leq 0$ ，所以权向量不变。

因为 $w_2 x_7^T \leq 0$ ，所以权向量不变。

因为 $w_2 x_8^T \leq 0$ ，所以权向量不变。

$$w_4 = (-1, -1, 6)$$

经过计算， $w_3 = (-1, -1, 6)$ 可以将所有样本分类正确，感知机算法收敛。

决策函数为：

$$g(\mathbf{x}) = -x_1 - x_2 + 6$$

当 $g(\mathbf{x}) > 0$ 时，判定为第一类样本；

当 $g(\mathbf{x}) \leq 0$ 时，判定为第二类样本。

题目：使用最小距离分离器，求决策函数。

数据：正类 4 个样本 $\{(1, 1), (1, 2), (2, 2), (1, 0)\}$ ；

负类 4 个样本 $\{(4, 5), (5, 6), (6, 7), (6, 6)\}$ ；

解:

类别均值作为原型向量:

$$\mathbf{m}_1 = \begin{pmatrix} 1.25 \\ 1.25 \end{pmatrix}$$

$$\mathbf{m}_2 = \begin{pmatrix} 5.25 \\ 6 \end{pmatrix}$$

决策准则:

$$h(\mathbf{x}) = \arg \min_{c_i} \| \mathbf{x} - \mathbf{m}_i \|$$

决策函数为：

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$$

$$\mathbf{w} = \mathbf{m}_1 - \mathbf{m}_2 = \begin{pmatrix} -4 \\ -4.75 \end{pmatrix}$$

$$b = -\frac{1}{2}(\mathbf{m}_1 - \mathbf{m}_2)^T (\mathbf{m}_1 + \mathbf{m}_2) = 30.21875$$

决策函数为：

$$g(\mathbf{x}) = -4x_1 - 4.75x_2 + 30.21875$$

当 $g(\mathbf{x}) > 0$ 时，判定为第一类样本；

当 $g(\mathbf{x}) \leq 0$ 时，判定为第二类样本。

题目：使用最小平方准则，求决策函数。

数据：正类 4 个样本 $\{(1, 1), (1, 2), (2, 2), (1, 0)\}$ ；

负类 4 个样本 $\{(4, 5), (5, 6), (6, 7), (6, 6)\}$ ；

解：

使用为逆矩阵法：

$$\mathbf{w} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

$$\mathbf{X}^T = \begin{bmatrix} 1 & 1 & 2 & 1 & 4 & 5 & 6 & 6 \\ 1 & 2 & 2 & 0 & 5 & 6 & 7 & 6 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\mathbf{y}^T = [1 \quad 1 \quad 1 \quad 1 \quad -1 \quad -1 \quad -1 \quad -1]$$

$$X^T X = \begin{bmatrix} 120 & 135 & 26 \\ 135 & 155 & 29 \\ 26 & 29 & 8 \end{bmatrix}$$

$$(X^T X)^{-1} = \begin{bmatrix} 0.4534 & -0.3705 & -0.1307 \\ -0.3705 & 0.3227 & 0.0341 \\ -0.1307 & 0.0341 & 0.4261 \end{bmatrix}$$

$$w = \begin{bmatrix} -0.2159 \\ -0.2045 \\ 1.4432 \end{bmatrix}$$

决策函数为：

$$g(\mathbf{x}) = -0.2159x_1 - 0.2045x_2 + 1.4432$$

当 $g(\mathbf{x}) > 0$ 时，判定为第一类样本；

当 $g(\mathbf{x}) \leq 0$ 时，判定为第二类样本。

题目：完成 Logistic 模型的算法过程说明，两分类问题。

数据：正类 4 个样本 $\{(1, 1), (1, 2), (2, 2), (1, 0)\}$ ；

负类 4 个样本 $\{(4, 5), (5, 6), (6, 7), (6, 6)\}$ ；

- (1) 给出目标函数表达式以及参数更新公式。
- (2) 给出迭代 1000 轮次后参数的值。

解 (1) : $\ln \frac{p(\omega_1 | \mathbf{x})}{p(\omega_0 | \mathbf{x})} = w_0 + w_1 x_1 + w_2 x_2$

$$p(\omega_1 | \mathbf{x}) + p(\omega_0 | \mathbf{x}) = 1$$

推出：

$$\begin{cases} p(\omega_0 | \mathbf{x}) = \frac{1}{1 + \exp(g(\mathbf{x}))} = \frac{1}{1 + \exp(w_0 + w_1 x_1 + w_2 x_2)} \\ p(\omega_1 | \mathbf{x}) = \frac{1}{1 + \exp(-g(\mathbf{x}))} = \frac{1}{1 + \exp(-(w_0 + w_1 x_1 + w_2 x_2))} \end{cases}$$

目标函数为：

$$\begin{aligned} L(\mathbf{w}) &= -\sum_{k=1}^{N_1} \ln(p(\omega_1 | \mathbf{x}_k^{(1)})) - \sum_{k=1}^{N_0} \ln(p(\omega_0 | \mathbf{x}_k^{(0)})) + C \\ &= -\sum_{k=1}^4 \ln\left(\frac{1}{1 + \exp(-g(\mathbf{x}_k^{(1)}))}\right) - \sum_{k=1}^4 \ln\left(\frac{1}{1 + \exp(g(\mathbf{x}_k^{(0)}))}\right) + C \end{aligned}$$

梯度为：

$$\frac{\partial L}{\partial \mathbf{w}} = \sum_{k=1}^4 \left(1 - \frac{1}{1 + \exp(-g(\mathbf{x}_k^{(1)}))}\right) \mathbf{x}_k^{(1)} - \sum_{k=1}^4 \left(1 - \frac{1}{1 + \exp(g(\mathbf{x}_k^{(0)}))}\right) \mathbf{x}_k^{(0)}$$

参数更新公式：

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \eta \frac{\partial L}{\partial \mathbf{w}}$$

解 (2):

使用梯度上升算法:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \eta \frac{\partial L}{\partial \mathbf{w}}, \text{ 学习率设为 } 1, \text{ 迭代 } 1000 \text{ 轮, 参数初始化为 } 0$$

向量解得:

$$w_0 = 17.79$$

$$w_1 = -1.57$$

$$w_2 = -3.82$$

决策函数为：

$$g(\mathbf{x}) = -1.57x_1 - 3.82x_2 + 17.79$$

当 $g(\mathbf{x}) > 0$ 时，判定为第一类样本；

当 $g(\mathbf{x}) \leq 0$ 时，判定为第二类样本。

题目：Fisher 鉴别分析求投影方向。

数据：正类 4 个样本 $\{(1, 1), (1, 2), (2, 2), (1, 0)\}$;

负类 4 个样本 $\{(4, 5), (5, 6), (6, 7), (6, 6)\}$;

解：第一类样本： $\{(1, 1), (1, 2), (2, 2), (1, 0)\}$

第二类样本： $\{(4, 5), (5, 6), (6, 7), (6, 6)\}$

$$\mu_1 = \begin{pmatrix} 1.25 \\ 1.25 \end{pmatrix} \quad \mu_2 = \begin{pmatrix} 5.25 \\ 6 \end{pmatrix}$$

$$\begin{aligned} S_w &= \sum_{x \in C_1} (x - \mu_1)(x - \mu_1)^T + \sum_{x \in C_2} (x - \mu_2)(x - \mu_2)^T \\ &= \begin{pmatrix} 0.75 & 0.75 \\ 0.75 & 2.75 \end{pmatrix} + \begin{pmatrix} 2.75 & 2 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} 3.5 & 2.75 \\ 2.75 & 4.75 \end{pmatrix} \end{aligned}$$

$$S_w^{-1} = \begin{pmatrix} 0.5241 & -0.3034 \\ -0.3034 & 0.3862 \end{pmatrix}$$

$$\omega = S_w^{-1}(\mu_1 - \mu_2) = \begin{pmatrix} -0.6551 \\ -0.6207 \end{pmatrix}$$

两类投影中心：

$$u_1 = \omega^T \mu_1 = -1.595 \qquad u_2 = \omega^T \mu_2 = -7.164$$

第一类样本： $\{-1.276, -1.897, -2.552, -0.655\}$

第二类样本： $\{-5.724, -7, -8.276, -7.655\}$

决策函数为：

$$g(\mathbf{x}) = \boldsymbol{\omega}^T \mathbf{x} - \frac{1}{2}(u_1 + u_2)$$

$$g(\mathbf{x}) = -0.6551x_1 - 0.6207x_2 + 4.379$$

当 $g(\mathbf{x}) > 0$ 时，判定为第一类样本；

当 $g(\mathbf{x}) \leq 0$ 时，判定为第二类样本。

总结： 数据： 正类 4 个样本 $\{(1, 1), (1, 2), (2, 2), (1, 0)\}$;

负类 4 个样本 $\{(4, 5), (5, 6), (6, 7), (6, 6)\}$;

各种线性分类器的决策函数如下：

线性分类器	决策函数
感知机	$g(\mathbf{x}) = -x_1 - x_2 + 6$
最小距离分类器	$g(\mathbf{x}) = -4x_1 - 4.75x_2 + 30.21875$
最小平方准则	$g(\mathbf{x}) = -0.2159x_1 - 0.2045x_2 + 1.4432$
Logistic 鉴别	$g(\mathbf{x}) = -1.57x_1 - 3.82x_2 + 17.79$
Fisher 鉴别	$g(\mathbf{x}) = -0.6551x_1 - 0.6207x_2 + 4.379$