课程作业

杨佳乐

2021 年 10 月 20 日

题目 1. 4.1

解答. 使用拉格朗日乘子法可以将其转换为标准特征值问题, 问题可表示为下面形式:

$$max_\alpha \qquad \alpha^T B\alpha$$

$$s.t. \qquad \alpha^T B\alpha = 1$$

转换成拉格朗日形式:

$$L(\alpha,\lambda) = \alpha^T B\alpha + \lambda(\alpha^T W\alpha - 1)$$

分别对 α, λ 求偏导:

$$\frac{\partial L(\alpha,\lambda)}{\partial \alpha} = 2B\alpha + 2\lambda W\alpha$$

$$\frac{\partial L(\alpha,\lambda)}{\partial \lambda} = \alpha^T W\alpha - 1$$

使 L 函数对 α 的偏导数等于 0，可以推导出:

$$-B\alpha = \lambda W\alpha$$

$$-W^{-1}B\alpha = \lambda\alpha$$

即为 $Ax = \lambda x$ 的标准形式。当 α 取 $-W^{-1}B$ 的特征向量，λ 取其特征值时，可以使目标函数最大。

题目 2. 4.2

解答. (a)

多元正态分布的概率密度函数可以成：

$$p_k(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T\Sigma^{-1}(x-\mu_k)}$$

对其取对数：

$$log(p_k(x)) = -\frac{1}{2}ln(|\Sigma|) - \frac{p}{2}ln(2\pi) - \frac{1}{2}(x-\mu_k)^T\Sigma^{-1}(x-\mu_k)$$

所以判别方程：

$$\delta_k(x) = log(p_k(x)) + ln(\pi_k)$$

以上判别式假设不同种类样本的协方差都为 Σ, 可以将常数项丢弃，得到简化后的判别式：

$$\delta_k(x) = -\frac{1}{2}(x-\mu_k)^T\Sigma^{-1}(x-\mu_k) + ln(\pi_k)$$

将第一项展开：

$$\delta_k(x) = -\frac{1}{2}(x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu_k - \mu_k^T\Sigma^{-1}x + \mu_k^T\Sigma^{-1}\mu_k) + ln(\pi_k)$$

$x^T\Sigma^{-1}x$ 在 k 为每一个值的判别式中都一样可以省略，因为协方差矩阵使对称的，所以 $x^T\Sigma^{-1}\mu_k = \mu_k^T\Sigma^{-1}x$，最后得到：

$$\delta_k(x) = x^T\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + ln(\pi_k)$$

假设 x 的标签是 2，则 $\delta_2(x) > \delta_1(x)$, 再将 $\pi_1 = \frac{N_1}{N}$， $\pi_2 = \frac{N_2}{N}$ 带入，即可推出题目给的不等式。

(b)

要使平方损失最小，需要满足 $\beta^* = (X^TX)^{-1}X^Ty$, 所以：

$$X^TX\beta^* = X^Ty \tag{1}$$

等式左边:

$$X^TX = \begin{bmatrix} 1 & \cdots & 1 & \cdots & 1 \\ x_1 & \cdots & x_{N_1} & \cdots & x_{N_1+N_2} \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ & x_{N_1} \\ \vdots & \vdots \\ & x_{N_1+N_2} \end{bmatrix}$$

$$= \begin{bmatrix} N & \sum_{i=1}^{N} x_i^T \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i x_i^T \end{bmatrix}$$

其中 $\sum_{i=1}^{N} x_i = \sum_{i=1}^{N_1} x_i + \sum_{i=N_1+1}^{N} x_i = N_1\mu_1 + N_2\mu_2$。同理可得 $\sum_{i=1}^{N} x_i^T = N_1\mu_1^T + N_2\mu_2^T$。对于 $(X^TX)_{2,2}$, 引入协方差矩阵:

$$\Sigma = \frac{1}{N-2}(\sum_{i=1}^{N_1}(x_i-\mu_1)(x_i-\mu_1)^T + \sum_{i=N_1+1}^{N}(x_i-\mu_2)(x_i-\mu_2)^T)$$

$$= \frac{1}{N-2}(\sum_{i=1}^{N_1}(x_i-\mu_1)(x_i^T-\mu_1^T) + \sum_{i=N_1+1}^{N}(x_i-\mu_2)(x_i^T-\mu_2^T))$$

$$= \frac{1}{N-2}(\sum_{i=1}^{N_1}(x_ix_i^T - x_i\mu_1^T - \mu_1x_i^T + \mu_1\mu_1^T) + \sum_{i=N_1+1}^{N}(x_ix_i^T - x_i\mu_2^T - \mu_2x_i^T + \mu_2\mu_2^T))$$

其中:

$$\sum_{i=1}^{N_1} x_i\mu_1^T = \sum_{i=1}^{N_1} \begin{bmatrix} x_{i1}\mu_{11} & x_{i1}\mu_{12} & \cdots & x_{i1}\mu_{1n} \\ \vdots & \ddots & & \vdots \\ x_{in}\mu_1 1 & \cdots & \cdots & x_{in}\mu 1n \end{bmatrix}$$

$$= \begin{bmatrix} N_1\mu_{11}^2 & N_1\mu_{11}\mu_{12} & \cdots & N_1\mu_{11}\mu_{1n} \\ \vdots & \ddots & & \vdots \\ N_1\mu_{1n}\mu_1 1 & \cdots & \cdots & N_1\mu_{1n}^2 \end{bmatrix}$$

$$= N_1\mu_1\mu_1^T$$

同理可得: $\sum_{i=1}^{N_1} \mu_1 x_i^T = N_1\mu_1\mu_1^T, \sum_{i=N_1+1}^{N} x_i\mu_2^T = N_2\mu_2\mu_2^T, \sum_{i=N_1+1}^{N} \mu_2 x_i^T = N_2\mu_2\mu_2^T$

再带入前面的方程可得：

$$\sum_{i=1}^{N} x_i x_i^T = (N-2)\Sigma + N_1\mu_1\mu_1^T + N_2\mu_2\mu_2^T$$

(1) 等式的右边，由题目所给条件，$1-N_1$ 是误分类点，其函数值是 $-\frac{N}{N_1}$,N_1-N（共 N_2 个点）是正确分类点，其函数值是 $\frac{N}{N_2}$：

$$X^T y = \begin{bmatrix} 1 & \cdots & 1 & 1 & \cdots & 1 \\ x_1 & \cdots & x_{N_1} & x_{N_1+1} & \cdots & x_N \end{bmatrix} \begin{bmatrix} -\frac{N}{N_1} \\ \vdots \\ -\frac{N}{N_1} \\ \frac{N}{N_2} \\ \vdots \\ \frac{N}{N_2} \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ -\frac{N}{N_1}\sum_{i=1}^{N_1} x_i + \frac{N}{N_2}\sum_{i=N_1+1}^{N} x_i \end{bmatrix} = \begin{bmatrix} 0 \\ N(\mu_2 - \mu_1) \end{bmatrix}$$

综上，（1）可以表示为：

$$\begin{bmatrix} N & N_1\mu_1^T + N_2\mu_2^T \\ N_1\mu_1 + N_2\mu_2 & (N-2)\Sigma + N_1\mu_1\mu_1^T + N_2\mu_2\mu_2^T \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ N(\mu_2 - \mu_1) \end{bmatrix}$$

它等价于两个方程：

$$N\beta_0 + (N_1\mu_1^T + N_2\mu_2^T)\beta = 0 \tag{2}$$

$$(N_1\mu_1 + N_2\mu_2)\beta_0 + ((N-2)\Sigma + N_1\mu_1\mu_1^T + N_2\mu_2\mu_2^T)\beta = N(\mu_2 - \mu_1) \tag{3}$$

由（2）可以推出 β 和 β_0 的关系，再带入（3）中可得：

$$left = ((N-2)\Sigma + N_1\mu_1\mu_1^T + N_2\mu_2\mu_2^T - \frac{N_1^2}{N}\mu_1\mu_1^T - \frac{2N_1N_2}{N}\mu_1\mu_2^T - \frac{N_2^2}{N}\mu_2\mu_2^T)\beta$$

$$= ((N-2)\Sigma + \frac{N_1}{N}(N-N_1)\mu_1\mu_1^T + \frac{N_2}{N}(N-N_2)\mu_2\mu_2^T - \frac{2N_1N_2}{N}\mu_1\mu_2^T)\beta$$

$$= ((N-2)\Sigma + \frac{N_1N_2}{N}(\mu_1-\mu_2)(\mu_1-\mu_2)^T)\beta$$

由题目条件可知 $\Sigma_B = \frac{N_1N_2}{N^2}(\mu_1-\mu_2)(\mu_1-\mu_2)^T$ 所以：

$$[(N-2)\Sigma + N\Sigma_B]\beta = N(\mu_2-\mu_1)$$

(C)

$$((N-2)\Sigma + \frac{N_1N_2}{N}(\mu_2-\mu_1)(\mu_2-\mu_1)^T)\beta = N(\mu_2-\mu_1)$$

令 $A = \frac{N_1N_2}{N}(\mu_2-\mu_1)^T\beta$,$A$ 是一个标量：

$$(N-2)\Sigma\beta = (N-A)(\mu_2-\mu_1)$$

所以：

$$\beta = \frac{N-A}{N-2}\Sigma^{-1}(\mu_2-\mu_1)$$

β 的方向和 $\Sigma^{-1}(\mu_2-\mu_1)$ 的方向一致。

题目 3. 4.3

解答.

题目 4. 4.6

解答. (a)

感知机可以表示为 $y_i\beta^Tx_i^* > 0$，两边同时除 $\|x_i^*\|$，可以得到：

$$y_i\beta^Tz_i > 0$$

在所有样本点 (x_i^*, y_i) 中，令 $m = min\{y_i\beta^Tx_i^*\}$。则 $y_i\beta^Tx_i^* \geq m$ 成立。两边同时除以 m, 令 $\beta_{sep} = \frac{1}{m}\beta$，即可推出题目要求不等式：

$$y_i\beta_{sep}^Tz_i \geq 1$$

(b)

更新超平面公式为 $\beta_{new} = \beta_{old} + y_i z_i$，两边同时减去 βsep 得到：

$$\beta_{new} - \beta_{sep} = \beta_{old} - \beta sep + y_i z_i$$

等式两边同时平方：

$$\|\beta_{new} - \beta_{sep}\|^2 = \|\beta_{old} - \beta sep\|^2 + y_i^2\|z_i\|^2 + 2y_i(\beta_{old} - \beta sep)^T z_i$$

其中 $y_i^2\|z_i\|^2 = 1$, 右边第三项中，因为 (y_i, x_i) 是误分类点，所以对应的 $y_i\beta_{old}^T z_i < 0$

又由（a）结论可得 $y_i\beta_{sep}^T z_i \geq 1$, 所以 $2y_i(\beta_{old} - \beta sep)^T z_i \leq -2$。最后可以推出：

$$\|\beta_{new} - \beta_{sep}\|^2 \leq \|\beta_{old} - \beta sep\|^2 + 1 - 2 = \|\beta_{old} - \beta sep\|^2 - 1$$

可以看出如果所有样本点可分，则最多 $\|\beta_{old} - \beta sep\|^2$ 步，算法可收敛。