机器学习作业4

181860152 周宇翔

1.[30pts] SVM with Weighted Penalty

(1)

将所有样本以及标签重新排序为$(\mathbf{x_1},y_1),...,(\mathbf{x_d},y_d),(\mathbf{x_{d+1}},y_{d+1}),...,(\mathbf{x_m},y_m)$

使得$\forall i \in \{1,2,\ldots,d\}, y_i = +1, \forall j \in \{d+1,\ldots,m\}, y_j = -1$

原本的SVM optimization problem转化为:

$$min_{\mathbf{w},b,\epsilon,\epsilon_i} \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{d}\epsilon_i + kC\sum_{i=d+1}^{m}\epsilon_i$$
$$s.t.\ y_i(\mathbf{w}^T\mathbf{x_i}+b) \geq 1-\epsilon_i$$
$$\epsilon_i \geq 0 \quad for \quad i=1,2,\ldots,d,d+1,\ldots,m$$

(2)

该式引入拉格朗日算子的拉格朗日函数为

$$\mathbf{L}(\mathbf{w},b,\alpha,\epsilon,\mu) = \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{d}\epsilon_i + kC\sum_{i=d+1}^{m}\epsilon_i + \sum_{i=1}^{m}\alpha_i(1-\epsilon_i-y_i(\mathbf{w}^T\mathbf{x_i}+b)) - \sum_{i=1}^{m}\mu_i\epsilon_i$$

令$\mathbf{L}$对$\mathbf{w},b,\epsilon_i$的偏导为0可得

$$\mathbf{w} = \sum_{i=1}^{m}\alpha_i y_i \mathbf{x}_i$$
$$0 = \sum_{i=1}^{m}\alpha_i y_i$$
$$\epsilon_i + \mu_i = C \quad for \quad i=1,\ldots,d$$
$$\epsilon_i + \mu_i = kC \quad for \quad i=d+1,\ldots,m$$

将上式代入拉格朗日函数可得对偶问题为

$$max_\alpha s\sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \mathbf{x}_i^T\mathbf{x_j}$$
$$s.t.\sum_{i=1}^{m}\alpha_i y_i = 0$$
$$0 \leq \alpha_i \leq C \quad i=1,\ldots,d$$
$$0 \leq \alpha_i \leq kC \quad i=d+1,\ldots,m$$

2.[35pts]Nearest Neignbor

(1)

$$Pr(d^* \leq t) = P(min_{1 \leq i \leq n}||\mathbf{x}_i|| \leq t)$$
$$= 1 - P(min_{1 \leq i \leq n}||\mathbf{x}_i|| > t)$$
$$= 1 - \prod_{i=1}^{n} P(||\mathbf{x}_i|| > t)$$
$$P(||\mathbf{x}|| > t) = 1 - \frac{\frac{4}{3}\pi t^3}{\frac{4}{3}\pi} = 1 - t^3$$
$$Pr(d^* \leq t) = 1 - (1 - t^3)^n$$

(2)

也即

$$P(||\mathbf{x} > t||) = 1 - \frac{\frac{(t\sqrt{\pi})^p}{\tau(p/2+1)}}{\frac{(\sqrt{\pi})^p}{\tau(p/2+1)}} = 1 - t^p$$

代回到(1)中可得$Pr(d^* \leq t) = 1 - (1 - t^p)^n$

(3)

也即$1 - (1 - t^p)^n = \frac{1}{2}$

$t = (1 - \frac{1}{2^{\frac{1}{n}}})^{\frac{1}{p}}$

3.[30pts] Principal Component Analysis

(1)

相似性:

两者都通过降维的方法来减少特征属性的数目从而避免维数灾难

不同处:

LDA是监督方法,要求样本有对应的标签,而PCA是无监督方法,不要求有标签.LDA降维的依据是降维后标签相同的样本之间应尽量接近,而,标签不同的样本之间应尽量远离.PCA降维的依据则是同一样本降维后的重构向量和它本身的距离应尽量小,且降维后的不同样本之间应该尽量分散

(2)

中心化:

$\frac{1}{3}\sum_{n=1}^{3} x_i = (0, \frac{4}{3})$

$(-2, 2) \rightarrow (-2, \frac{2}{3}), (0, 0) \rightarrow (0, -\frac{4}{3}), (2, 2) \rightarrow (2, \frac{2}{3})$

协方差矩阵:

$$\begin{bmatrix} -2 & 0 & 2 \\ \frac{2}{3} & -\frac{4}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} -2 & \frac{2}{3} \\ 0 & -\frac{4}{3} \\ 2 & \frac{2}{3} \end{bmatrix} = \begin{bmatrix} 8 & 0 \\ 0 & 8 \end{bmatrix}$$

特征值分解:

很明显特征值为8,特征向量为(1,0) 或 (0,1)

投影矩阵

$$W^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

(3)

$$d' = 1, W^* = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

投影后的坐标分别为 -2, 0 ,2