Introduction to Machine Learning
Homework 4

吴紫航 171860659

1 [30pts] SVM with Weighted Penalty

Consider the standard SVM optimization problem as follows (i.e., formula (6.35)in book),

$$
\begin{aligned}
\min_{\mathbf{w},b,\xi_i} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{m}\xi_i \\
\text{s.t.} \quad & y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) \geq 1 - \xi_i \\
& \xi_i \geq 0, i = 1, 2, \cdots, m.
\end{aligned}
\tag{1.1}
$$

Note that in (1.1), for positive and negative examples, the "penalty" of the classification error in the objective function is the same. In the real scenario, the price of "punishment" is different for misclassifying positive and negative examples. For example, considering cancer diagnosis, misclassifying a person who actually has cancer as a healthy person, and misclassifying a healthy person as having cancer, the wrong influence and the cost should not be considered equivalent.

Now, we want to apply $k > 0$ to the "penalty" of the examples that were split in the positive case for the examples with negative classification results (i.e., false positive). For such scenario,

(1) [15pts] Please give the corresponding SVM optimization problem;

(2) [15pts] Please give the corresponding dual problem and detailed derivation steps, especially such as KKT conditions.

(1)解:

由题意FP样本的惩罚量是原先的k倍，FN样本的惩罚量维持不变。不妨假

1

设$y_1, y_2, \cdots, y_m = 1$，且$y_{m+1}, y_{m+2}, \cdots, y_{m+n} = -1$。即m个正例和n个负例。那么相应的SVM优化问题为：

$$\min_{\mathbf{w}, b, \xi_i, \xi_j} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C_1 \sum_{i=1}^{m} \xi_i + C_2 \sum_{j=m+1}^{m+n} \xi_j$$

$$\begin{aligned}
\text{s.t.} \quad & y_p(\mathbf{w}^{\mathrm{T}}\mathbf{x}_p + b) \geq 1 - \xi_p \\
& \xi_p \geq 0, p = 1, 2, \cdots, m, m+1, \cdots, m+n; \\
& C_1 = kC_2 = kC; \\
& y_1, y_2, \cdots, y_m = +1; \\
& y_{m+1}, y_{m+2}, \cdots, y_{m+n} = -1.
\end{aligned} \tag{1.2}$$

(2)解：

由拉格朗日乘子法我们得到(1)中优化问题的拉格朗日函数

$$\begin{aligned}
L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = & \frac{1}{2}\|\boldsymbol{w}\|^2 + C_1 \sum_{i=1}^{m} \xi_i + C_2 \sum_{j=m+1}^{m+n} \xi_j \\
& + \sum_{p=1}^{m+n} \alpha_p(1 - \xi_p - y_p(\mathbf{w}^{\mathrm{T}}\mathbf{x}_p + b)) + \sum_{p=1}^{m+n} \beta_p(-\xi_p)
\end{aligned} \tag{1.3}$$

我们通过构造上述拉格朗日函数，把带有不等式约束的优化转化为无约束优化，且引入了松弛变量α和β，需要满足KKT条件如下

$$\begin{aligned}
\alpha_p \geq 0, \quad \beta_p \geq 0, \\
1 - \xi_p - y_p(\mathbf{w}^{\mathrm{T}}\mathbf{x}_p + b) \leq 0, \\
-\xi_p \leq 0, \\
\alpha_p(1 - \xi_p - y_p(\mathbf{w}^{\mathrm{T}}\mathbf{x}_p + b)) = 0, \\
\beta_p(-\xi_p) = 0.
\end{aligned} \tag{1.4}$$

即原优化问题转化为

$$\min_{\boldsymbol{w}, b, \boldsymbol{\xi}} \max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \tag{1.5}$$

对偶问题为

$$\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \min_{\boldsymbol{w},b,\boldsymbol{\xi}} L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \tag{1.6}$$

令 $L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ 对 $\boldsymbol{w}, b, \xi_p$ 求偏导为零得

$$\boldsymbol{w} = \sum_{p=1}^{m+n} \alpha_p y_p \boldsymbol{x}_p \tag{1.7}$$

$$0 = \sum_{p=1}^{m+n} \alpha_p y_p \tag{1.8}$$

$$C_1 = \alpha_i + \beta_i, \ i = 1, 2, \cdots, m \tag{1.9}$$

$$C_2 = \alpha_j + \beta_j, \ j = m+1, m+2, \cdots, m+n \tag{1.10}$$

将(1.7)-(1.10)代入拉格朗日函数(1.3)，即得到(1.2)优化问题的对偶问题

$$
\begin{aligned}
\max_{\boldsymbol{\alpha}} \quad & \sum_{p=1}^{m+n} \alpha_p - \frac{1}{2} \sum_{p=1}^{m+n} \sum_{q=1}^{m+n} \alpha_p \alpha_q y_p y_q \boldsymbol{x}_p^T \boldsymbol{x}_q \\
\text{s.t.} \quad & \sum_{p=1}^{m+n} \alpha_p y_p = 0 \\
& 0 \le \alpha_i \le kC, \ i = 1, 2, \cdots, m \\
& 0 \le \alpha_j \le C, \ j = m+1, m+2, \cdots, m+n
\end{aligned}
\tag{1.11}
$$

2   [35pts] Nearest Neighbor

Let $\mathcal{D} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be a set of instances sampled completely at random from a p-dimensional unit ball B centered at the origin, i.e.,

$$B = \left\{ \mathbf{x} : \|\mathbf{x}\|^2 \leq 1 \right\} \subset \mathbb{R}^p. \tag{2.1}$$

Here, $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ and $\langle \cdot, \cdot \rangle$ indicates the dot product of two vectors.

In this assignment, we consider to find the nearest neighbor for the origin. That is, we define the shortest distance between the origin and $\mathcal{D}$ as follows,

$$d^* := \min_{1 \leq i \leq n} \|\mathbf{x}_i\|. \tag{2.2}$$

It can be seen that d^* is a random variable since $\mathbf{x}_i, \forall 1 \leq i \leq n$ are sampled completely at random.

(1) [10pts] Assume $p = 3$ and $t \in [0, 1]$, calculate $\Pr(d^* \leq t)$, i.e., the cumulative distribution function (CDF) of random variable d^*.

解：设$d^\star$的累计函数为$F_D(t) = Pr(d^\star \leq t)$.

由于x_i是在单位球体内的，故d*的范围是[0,1]。

故当$t < 0, F_D(t) = 0$; 当$t > 1, F_D(t) = 1$;

当$0 \leq t \leq 1$时，$F_D(t) = Pr(d^\star \leq t) = 1 - Pr(d^\star > t) = 1 - \prod_{i=1}^{n} Pr(\|\boldsymbol{x}_i\| > t) = 1 - \prod_{i=1}^{n} \left(\frac{\frac{4\pi 1^3}{3} - \frac{4\pi t^3}{3}}{\frac{4\pi 1^3}{3}} \right) = 1 - \prod_{i=1}^{n} (1 - t^3) = 1 - (1 - t^3)^n$

因此$d^\star$的分布函数$F_D(t)$为

$$F_D(t) = \begin{cases} 0 & t < 0; \\ 1 - (1 - t^3)^n & 0 \leq t \leq 1; \\ 1 & t > 1. \end{cases} \tag{2.3}$$

(2) [15pts] Show the general formula of CDF of random variable d^* for $p \in \{1, 2, 3, \ldots\}$. You may need to use the volume formula of sphere with radius equals r,

$$V_p(r) = \frac{(r\sqrt{\pi})^p}{\Gamma(p/2 + 1)}. \tag{2.4}$$

Here, $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$, and $\Gamma(x+1) = x\Gamma(x), \forall x > 0$. For $n \in \mathbb{N}^*$, $\Gamma(n+1) = n!$.

解：对于$t < 0$和$t > 1$的情况，分布函数值还是0和1，不变。
当$0 \leq t \leq 1$时，$F_D(t) = Pr(d^\star \leq t) = 1 - Pr(d^\star > t) = 1 - \prod_{i=1}^{n} Pr(\|\boldsymbol{x}_i\| > t) = 1 - \prod_{i=1}^{n}(\frac{V_p(1) - V_p(t)}{V_p(1)}) = 1 - \prod_{i=1}^{n}(\frac{1^p - t^p}{1^p}) = 1 - (1 - t^p)^n$
因此$d^\star$的分布函数$F_D(t)$为

$$F_D(t) = \begin{cases} 0 & t < 0; \\ 1 - (1 - t^p)^n & 0 \leq t \leq 1; \quad p \in \{1, 2, 3, \dots\}. \\ 1 & t > 1. \end{cases} \qquad (2.5)$$

(3) [10pts] Calculate the median of the value of random variable d^*, i.e., calculate the value of t that satisfies $\Pr(d^* \leq t) = \frac{1}{2}$.

解：
由$Pr(d^\star \leq t) = F_D(t) = \frac{1}{2}$, 则$1 - (1 - t^p)^n = \frac{1}{2}$,
即$t = (1 - (0.5)^{\frac{1}{n}})^{\frac{1}{p}}$为中位数点

3 [30pts] Principal Component Analysis

(1) [10 pts] Please describe describe the similarities and differences between PCA and LDA.

解：

相同点：

(a)都是机器学习中对数据进行降维处理的方法

(b)都利用到矩阵特征分解的思想转化为优化问题

(c)都假设数据满足正态分布

不同点：

(a)LDA是监督学习，PCA是无监督学习

(b)LDA选择类内方差小,类间方差大的超平面;PCA选择方差大的超平面

(c)LDA除了降维，还可以用来分类

(d)LDA降维和类个数C有关，可以降维至1到C-1维;PCA降维没有这个限制

(e)PCA投影的坐标系是正交的，而LDA不保证(一般都不正交)

注：本题参考了如下链接

[1]机器学习降维方法:PCA和LDA的区别

https://blog.csdn.net/sinat_30353259/article/details/81569550

[2]PCA和LDA

https://blog.csdn.net/hlang8160/article/details/78768260

(2) [10 pts] Consider 3 data points in the 2-d space: (-2, 2), (0, 0), (2, 2), What is the first principal component? (Maybe you don't really need to solve any SVD or eigenproblem to see this.)

解：对于坐标$(-2,2),(0,0),(2,2)$,首先得到两个向量$\boldsymbol{X} = (-2,0,2)$, $\boldsymbol{Y} = (2,0,2)$. 再进行中心化得到$\boldsymbol{X} = (-2,0,2)$, $\boldsymbol{Y} = (\frac{2}{3}, -\frac{4}{3}, \frac{2}{3})$

接着得到协方差矩阵为

$$\boldsymbol{Q} = \left(\begin{array}{cc} Var(\boldsymbol{X}) & Cov(\boldsymbol{X},\boldsymbol{Y}) \\ Cov(\boldsymbol{X},\boldsymbol{Y}) & Var(\boldsymbol{Y}) \end{array} \right) = \frac{1}{3} \left(\begin{array}{cc} \boldsymbol{X}^T\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{Y} \\ \boldsymbol{X}^T\boldsymbol{Y} & \boldsymbol{Y}^T\boldsymbol{Y} \end{array} \right) \qquad (3.1)$$

即

$$\boldsymbol{Q} = \frac{1}{3} \begin{pmatrix} 8 & 0 \\ 0 & \frac{8}{3} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{8}{3} & 0 \\ 0 & \frac{8}{9} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad (3.2)$$

因为$\frac{8}{3} > \frac{8}{9}$，所以第一个主元是$x$轴,方向为$\boldsymbol{e} = (1,0)$.

(3) [10 pts] If we projected the data into 1-d subspace, what are their new coordinates?

解：把点投影到一维，即第二问求得的主元x轴，坐标从二维的$(-2,2)$、$(0,0)$、$(2,2)$变为一维的-2、0、2