# Machine Learning Assignment 6 (SVM)

Due: 2, June

## 1. [35pts] Support Vector Machine

(1) Recall that the soft margin support vector machine solves the problem:

$$min \quad \frac{1}{2}w^{\top}w + C\sum_{i} \varepsilon_i$$

$$\text{s.t.} \;\; y_i(w^{\top}x_i + b) \geq 1 - \varepsilon_i, \quad \varepsilon_i \geq 0.$$

a) [10pts] Derive its dual problem using the method of Lagrange multipliers.

b) [10pts] Further simplify the dual problem when at its saddle point to prove

$$\max_{\alpha} \sum_{i} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i\, \alpha_j y_i y_j x_i^{\top} x_j$$

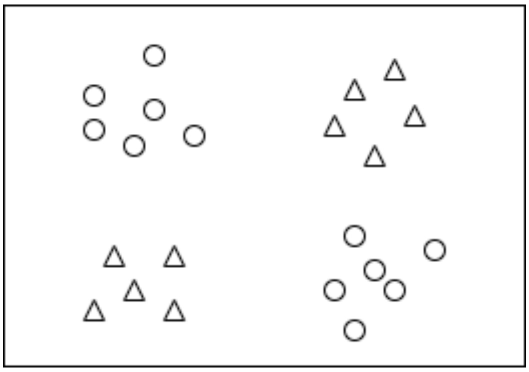
$$\text{s.t.} \;\; C \geq \alpha_i \geq 0, \quad \sum_{i} \alpha_i\, y_i = 0,$$

is equivalent to the primal problem.

(2) [15pts] Given the XOR sample points as below, we train an SVM with a quadratic kernel,

i.e. our kernel function is a polynomial kernel of degree 2: $\kappa(x_i, x_j) = \left(x_i^T x_j\right)^d, d = 2.$

(a) [5pts] what is the corresponding mapping function $\phi(x)$?



(b) [5pts] Use the following code to generate XOR data, and according to the answer of (a), map the data with $\phi(x)$ to see if it can be linearly separable.

(c) [5pts] Could we get a reasonable model with hard margin? If yes, draw the decision boundary in the figure (original feature space), otherwise state reasons.

```
import numpy as np
import matplotlib.pyplot as plt
#创建数据
X_xor = np.random.randn(40,2)
y_xor = np.logical_xor(X_xor[:,0]>0, X_xor[:,1]>0)
y_xor = np.where(y_xor, 1, -1)
#绘制散点图
plt.scatter(x=X_xor[y_xor==1,0]),   # 横坐标
            y=X_xor[y_xor==1,1]),   # 纵坐标
            color='g', marker='x', label='1')
plt.scatter(x=X_xor[y_xor==-1,0]),
            y=X_xor[y_xor==-1,1]),
            color='b', marker='o', label='-1')
plt.legend() #显示图例
plt.show()
```

## 2. [30pts] Kernel Functions

(1) **[15 pts]** 对于 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$，考虑函数 $\kappa(x, y) = \tanh(a\boldsymbol{x}^\top \boldsymbol{y} + b)$，其中 $a, b$ 是任意实数。试说明 $a \geq 0, b \geq 0$ 是 $\kappa$ 为核函数的必要条件。

(2) **[15 pts]** 考虑 $\mathbb{R}^N$ 上的函数 $\kappa(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}^\top \boldsymbol{y} + c)^d$，其中 $c$ 是任意实数，$d, N$ 是任意正整数。试分析函数 $\kappa$ 何时是核函数，何时不是核函数，并说明理由。

说明：该核函数是多项式核的更一般的形式。

（第（3）小题是 extra 部分，可选）

(3) [**10 pts**] 当上一小问中的函数是核函数时，考虑 $d = 2$ 的情况，此时 $\kappa$ 将 $N$ 维数据映射到了什么空间中？具体的映射函数是什么？更一般的，对 $d$ 不加限制时，$\kappa$ 将 $N$ 维数据映射到了什么空间中？(本小问的最后一问可以只写结果)

（可中文作答）

## 3. [35 pts] Kernel Methods

请给出 kernel PCA 的推导过程。

（可中文作答）

## 4. [extra, 30pts] Surrogate Function in SVM & Bayesian Optimal Classifier

（本题可选）

在软间隔支持向量机问题中，我们的优化目标为

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m} \ell_{0/1}\left(y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b\right) - 1\right). \tag{1}$$

然而 $\ell_{0/1}$ 数学性质不太好，它非凸、非连续，使得式 (1) 难以求解。实践中我们通常会将其替换为“替代损失”，替代损失一般是连续的凸函数，且为 $\ell_{0/1}$ 的上界，比如 hinge 损失，指数损失，对率损失。下面我们证明在一定的条件下，这样的替换可以保证最优解不变。

我们考虑实值函数 $h:\mathcal{X}\to\mathbb{R}$ 构成的假设空间，其对应的二分类器 $f_h:\mathcal{X}\to\{+1,-1\}$ 为

$$f_h(x) = \begin{cases} +1 & \text{if } h(x)\geq 0 \\ -1 & \text{if } h(x) < 0 \end{cases}$$

$h$ 的期望损失为 $R(h) = \mathbb{E}_{(x,y)\sim\mathcal{D}}\left[I_{f_h(x)\neq y}\right]$，其中 $I$ 为指示函数。设 $\eta(x) = \mathbb{P}(y=+1|x)$，则贝叶斯最优分类器当 $\eta(x)\geq\frac{1}{2}$ 时输出 1，否则输出 $-1$。因此可以定义贝叶斯得分 $h^*(x) = \eta(x) - \frac{1}{2}$ 和贝叶斯误差 $R^* = R(h^*)$。

设 $\Phi:\mathbb{R}\to\mathbb{R}$ 为非减的凸函数且满足 $\forall u\in\mathbb{R}, 1_{u\leq 0}\leq\Phi(-u)$。对于样本 $(x,y)$，定义函数 $h$ 在该样本的 $\Phi$-损失为 $\Phi(-yh(x))$，则 $h$ 的期望损失为 $\mathcal{L}_\Phi(h) = \underset{(x,y)\sim\mathcal{D}}{\mathbb{E}}[\Phi(-yh(x))]$。定义 $L_\Phi(x,u) = \eta(x)\Phi(-u) + (1-\eta(x))\Phi(u)$，设 $h_\Phi^*(x) = \underset{u\in[-\infty,+\infty]}{\operatorname{argmin}} L_\Phi(x,u)$，$\mathcal{L}_\Phi^* = \mathcal{L}_\Phi(h_\Phi^*(x))$。

我们考虑如下定理的证明：

若对于 $\Phi$，存在 $s\geq 1$ 和 $c>0$ 满足对 $\forall x\in\mathcal{X}$ 有

$$|h^*(x)|^s = \left|\eta(x) - \frac{1}{2}\right|^s \leq c^s\left[L_\Phi(x,0) - L_\Phi\left(x, h_\Phi^*(x)\right)\right] \tag{2}$$

则对于任何假设 $h$，有如下不等式成立

$$R(h) - R^* \leq 2c\left[\mathcal{L}_\Phi(h) - \mathcal{L}_\Phi^*\right]^{\frac{1}{s}} \tag{3}$$

(1) **[5 pts]** 请证明

$$\Phi\left(-2h^*(x)h(x)\right) \leq L_\Phi(x, h(x)) \tag{4}$$

(2) **[10 pts]** 请证明

$$R(h) - R^* = 2\underset{x\sim\mathcal{D}_x}{\mathbb{E}}\left[|h^*(x)|\, 1_{h(x)h^*(x)\leq 0}\right] \tag{5}$$

提示：先证明

$$R(h) = \underset{x\sim\mathcal{D}_x}{\mathbb{E}}\left[2h^*(x)1_{h(x)<0} + (1-\eta(x))\right]$$

(3) **[10 pts]** 利用式 (4) 和式 (5) 完成定理的证明。

(4) **[5 pts]** 请验证对于 Hinge 损失 $\Phi(u) = \max(0, 1+u)$，有 $s=1, c=\frac{1}{2}$。

## 5. Kernel Methods

推导 kernel LDA

（本题选做，自行作答，不用提交, 答案参考南瓜书对应章节）