Machine Learning Assignment 4 (SVM Part)

Due: TBD, submit with the part of "Bayesian Classifiers" later

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

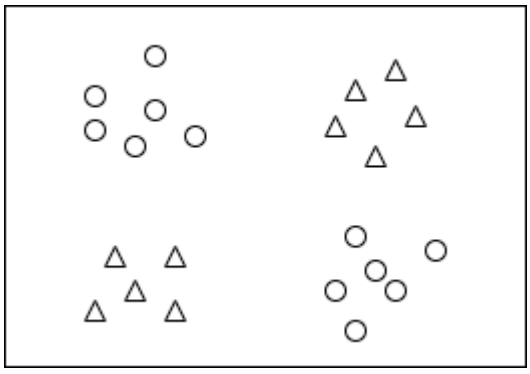
$$\max_{\alpha} \sum_{i}\alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\,\alpha_j y_i y_j x_i^{\top}x_j$$

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

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 (after feature mapping)? 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()
```

Solution

2. Kernel Methods [必做题，不提交不批改，可参考教材核对答案]

请给出 kernel PCA 的推导过程。

（可中文作答）

Solution

参见课本 10.4 节，或南瓜书对应章节。

3. Kernel Functions [选做题，不提交不批改，后续公布答案]

注：其中第(3)小题可以帮助深入理解核函数与特征映射函数之间的关系

(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 κ 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 κ 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。

(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 κ 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，κ 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）

4. Kernel Methods [选做题，不提交不批改，参考南瓜书核对答案]

推导 kernel LDA