

第六次作业参考答案

Exercise 6.1

线性SVM和一般线性分类器的区别主要是:

- A. 是否进行了空间映射
- B. 是否确保间隔最大化**
- C. 是否能处理线性不可分问题
- D. 训练误差通常较低

解析: 间隔最大化是SVM的特点。

Exercise 6.2

在SVM中, margin的含义是()

- A. 差额
- B. 损失误差
- C. 间隔**
- D. 幅度

解析: margin的含义是间隔。

Exercise 6.3

已知正例点 $x_1 = (1, 2)^T, x_2 = (2, 3)^T, x_3 = (3, 3)^T$, 负例点 $x_4 = (2, 1)^T, x_5 = (3, 2)^T$, 试求最大间隔分离超平面和分类决策函数, 并在图上画出分离超平面、间隔边界及支持向量。

解析:

原始形式

$$\min \frac{1}{2} \|w_1^2 + w_2^2\|$$

$$\text{s.t. } w_1 + 2w_2 + b \geq 1 \dots (1)$$

$$2w_1 + 3w_2 + b \geq 1 \dots (2)$$

$$3w_1 + 3w_2 + b \geq 1 \dots (3)$$

$$-2w_1 - w_2 - b \geq 1 \dots (4)$$

$$-3w_1 - 2w_2 - b \geq 1 \dots (5)$$

化简一下有

$$(1) + (4) : -w_1 + w_2 \geq 2$$

$$(1) + (5) : -2w_1 \geq 2$$

$$(2) + (4) : 2w_2 \geq 2$$

$$(3) + (4) : w_1 + 2w_2 \geq 2$$

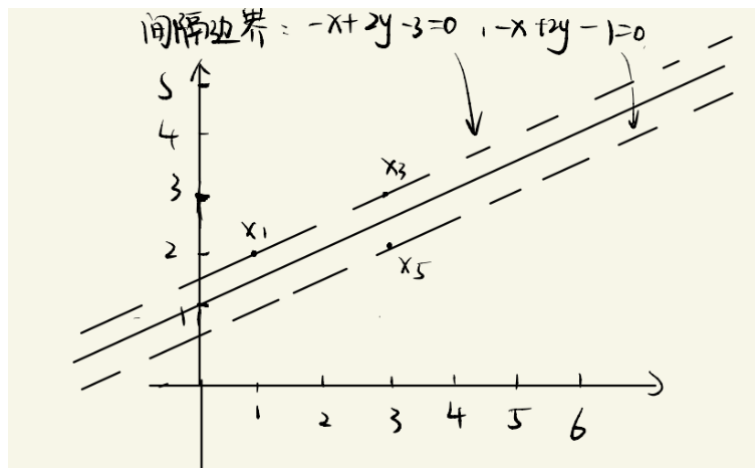
$$(3) + (5) : w_2 \geq 2$$

得到的5个方程, 使用高中知识数学规划, 画一下关于 w_1, w_2 坐标系, 发现 $w_1^2 + w_2^2$ 最小是在 $w_1 = -1, w_2 = 2$

将这个值带入原来的 (1) - (5) 方程, 可以得到 $b = -2$

则其最大间隔分离超平面为: $-x_1 + 2x_2 - 2 = 0$

其分类决策函数为: $y = \text{sign}(-x_1 + 2x_2 - 2)$



Exercise 6.4

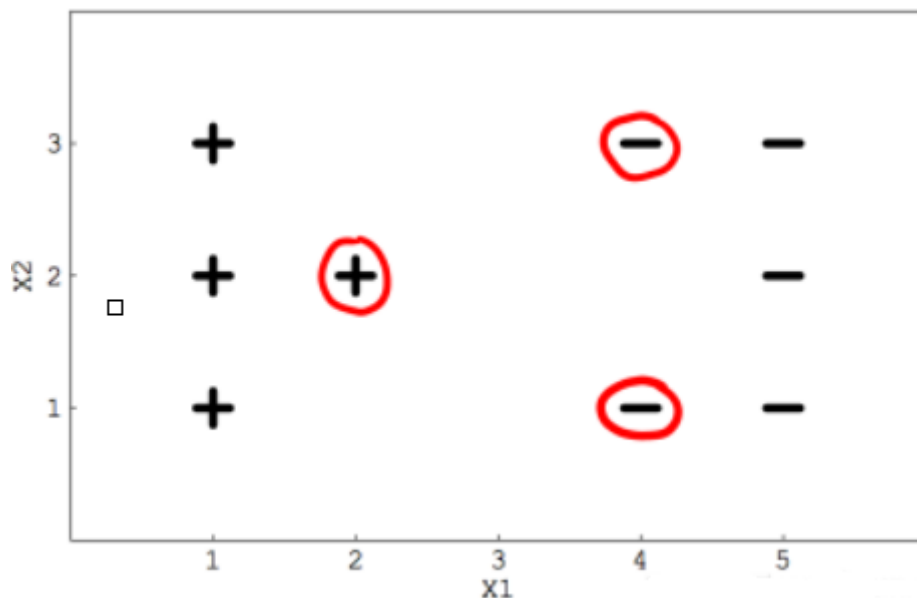
假设你训练SVM后，得到一个线性决策边界，你认为该模型欠拟合，在下次迭代训练模型时，应该考虑：

- A、增加训练数据
- B、减少训练数据
- C、计算更多变量**
- D、减少特征

解析：由于是欠拟合，考虑创造更多特征带入模型训练

Exercise 6.5

假定你用一个线性SVM分类器求解二类分类问题，如下图所示，这些用红色圆圈起来的点表示支持向量，据此回答问题1和2：



1.如果移除这些圈起来的数据，决策边界（即分离超平面）是否会发生改变？

- A. Yes **B. No**

解析：从数据的分布来看，移除那三个数据，决策边界不会受到影响。

2.如果将数据中除圈起来的三个点以外的其他数据全部移除，那么决策边界是否会改变？

- A. True **B. False**

解析：决策边界只会被支持向量影响，跟其他点无关。

Exercise 6.6

Consider two examples $(z_1; +1)$ and $(z_2; -1)$ with $z_1 = z$ and $z_2 = -z$. What is the Lagrange function $L(b; w; \lambda)$ of hard-margin SVM?

A $\frac{1}{2}w^T w + \lambda_1(1 + w^T z + b) + \lambda_2(1 + w^T z + b)$

B $\frac{1}{2}w^T w + \lambda_1(1 - w^T z - b) + \lambda_2(1 - w^T z + b)$

C $\frac{1}{2}w^T w + \lambda_1(1 + w^T z + b) + \lambda_2(1 + w^T z - b)$

D $\frac{1}{2}w^T w + \lambda_1(1 - w^T z - b) + \lambda_2(1 - w^T z - b)$

解析：

By definition,

$$\mathcal{L}(b, \mathbf{w}, \alpha) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \alpha_1(1 - y_1(\mathbf{w}^T \mathbf{z}_1 + b)) + \alpha_2(1 - y_2(\mathbf{w}^T \mathbf{z}_2 + b))$$

with $(\mathbf{z}_1, y_1) = (\mathbf{z}, +1)$ and $(\mathbf{z}_2, y_2) = (-\mathbf{z}, -1)$.

Exercise 6.7

For a single variable w , consider minimizing $\frac{1}{2}w^2$ subject to two linear constraints $w \geq 1$ and $w \leq 3$. We know that the Lagrange function $L(w, \alpha) = \frac{1}{2}w^2 + \lambda_1(1 - w) + \lambda_2(w - 3)$. Which of the following equations that contain λ are among the KKT conditions of the optimization problem?

A $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$

B $w = \lambda_1 - \lambda_2$

C $\alpha_1(1 - w) = 0$ and $\alpha_2(w - 3) = 0$

D all of the above

解析：根据KKT可得上述ABC结论。其中对偶问题可行，要求选项A成立。互补松弛条件，要求选项C成立。根据稳定性条件，对 $L(w, \lambda)$ 求导可知选项B成立。

Exercise 6.8

Consider two examples $(z_1; +1)$ and $(z_2; -1)$ with $z_1 = z$ and $z_2 = -z$. After solving the dual problem of hard-margin SVM, assume that the optimal λ_1 and λ_2 are both strictly positive. What is the optimal b ?

A -1

B 0

C 1

D not certain with the descriptions above

解析：根据条件可知， z_1 和 z_2 都是支持向量，则他们都在最大间隔的边界上。

故 $wz + b = 1 = -(-wz + b)$, 计算可得 $b = 0$

Exercise 6.9

Consider applying dual hard-margin SVM on $N = 5566$ examples and getting 1126 SVs. Which of the following can be the number of examples that are on the fat boundary—that is, SV candidates?

- A 0
- B 1024
- C 1234**
- D 9999

解析：在最大间隔边界上的样本点个数会多余支持向量，并不多于样本总数。

Exercise 6.10

对于在原空间中线性不可分问题，支持向量机（ ）。

- A. 在原空间中寻找非线性函数的划分数据
- B. 无法处理
- C. 在原空间中寻找线性函数划分数据
- D. 将数据映射到核空间中**

解析：使用核函数将原数据转换到高维空间中

Exercise 6.11 (多选题)

支持向量机有哪些常用的核函数()。

- A. 高斯核**
- B. 拉普拉斯核**
- C. 线性核**
- D. 多项式核**

解析：上述都是常用的核函数。

Exercise 6.12 (多选题)

以下关于支持向量机的说法正确的是()。

- A. SVM适用于大规模数据集
- B. SVM分类思想就是将分类面之间的间隔最小化
- C. SVM方法简单，鲁棒性较好**
- D. SVM分类面取决于支持向量**

解析：SVM适用于中等规模的数据集，其分类思想是讲分类面的间隔最大化

Exercise 6.13

关于支持向量机中硬间隔和软间隔的说法错误的是（ ）。

- A. 硬间隔有利于消除模型的过拟合**
- B. 硬间隔要求所有数据分类完全准确，不允许出现错误
- C. 软间隔允许一定的样本分类错误
- D. 软间隔有利于获取更大的分类间隔

解析：硬间隔的最小化函数中没有对错误的冗余，相比来说易过拟合。

Exercise 6.14

SVM 原理描述不正确的是（ ）。

- A. 当训练样本线性可分时，通过硬间隔最大化，学习一个线性分类器，即线性可分支持向量机
- B. 当训练数据近似线性可分时，引入松弛变量，通过软间隔最大化，学习一个线性分类器，即线性支持向量机
- C. SVM 的基本模型是在特征空间中寻找间隔最小化的分离超平面的线性分类器**
- D. 当训练数据线性不可分时，通过使用核技巧及软间隔最大化，学习非线性支持向量机

解析： SVM是寻找间隔最大化的线性分类器

Exercise 6.15

SVM普遍使用的准则描述不正确的是：（ ）(n 为特征数， m 为训练样本数。)

- A. 如果相较于 m 而言， n 要大许多，即训练集数据量不够支持我们训练一个复杂的非线性模型，我们选用逻辑回归模型或者不带核函数的支持向量机。
- B. 如果 n 较小，而且 m 大小中等，例如 n 在1 – 1000 之间，而 m 在10 – 10000之间，使用高斯核函数的支持向量机。
- C. 支持向量机理论上不能处理太多的特征。**
- D. 如果 n 较小，而 m 较大，例如 n 在1 – 1000之间，而 m 大于50000，则使用支持向量机会非常慢，解决方案是创造、增加更多的特征，然后使用逻辑回归或不带核函数的支持向量机。

解析： 在支持向量机中引入核技巧可以处理多特征，会学习到一个复杂的决策边界

Exercise 6.16

假设你用RBF核训练了一个SVM分类器，看起来似乎对训练集欠拟合，你应该提升还是降低 γ (gamma) ？C呢？

解析： 如果一个使用RBF核训练的支持向量机对欠拟合训练集，可能是由于过度正则化导致的。你需要提升gamma或C（或同时提升二者）来降低正则化