# Machine Learning Homework 4.1 & 4.2

专业：软件工程　　姓名：沈金龙　　学号：18214806

1、 Homework 4.1 SVM：

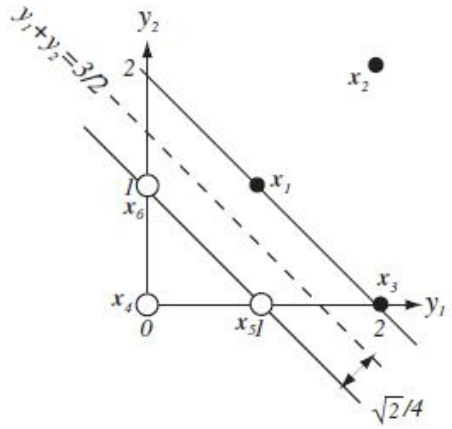
(a):

题目给出了包含两种分类的 6 个点，如下图所示：

| category | $x_1$ | $x_2$ |
|---|---|---|
| $\omega_1$ | 1 | 1 |
| $\omega_1$ | 2 | 2 |
| $\omega_1$ | 2 | 0 |
| $\omega_2$ | 0 | 0 |
| $\omega_2$ | 1 | 0 |
| $\omega_2$ | 0 | 1 |

其中，$w_1 : z_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$... 其余同理。

考虑 $z_1 = z_2 = z_3 = -1$ ，则其余为 +1，作图如下所示：




最优超平面为：$y_1 + y_2 = 3/2$

最优间隔为点到超平面的最近距离，为 $\sqrt{2}/4$

(b):

支持向量是边界上面的样本，即距离最优超平面最短的样本，故为：$z_1, z_3, z_5, z_6$

(c):

待定系数 a 的值为：$(3 \quad -2 \quad -2)^t$

2、 Homework 4.2 Neural Network：

(a):

Let $z^{h_i}$ and $z^o$ be the input to the sigmoid function at the hidden and output layers:

$$\frac{\partial l}{\partial w_{1,2}^{[1]}} = \frac{\partial l}{\partial o}\frac{\partial o}{\partial z^o}\frac{\partial z^o}{\partial h_2}\frac{\partial h_2}{\partial z^{h_2}}\frac{\partial z^{h_2}}{\partial w_{1,2}^{[1]}}$$

$$= \frac{1}{m}\sum_i^m 2o^{(i)}o^{(i)}(1\text{-}o^{(i)})w_2^{[2]}h_2(1-h_2)x_1^{(i)}$$

Here, we used the fact that for the sigmoid function:

$$\sigma(z) = \frac{1}{1+e^{(-z)}}$$

It's derivative w.s.t. $z$ is:

$$\frac{\partial \sigma}{\partial z} = \sigma(z)(1-\sigma(z))$$

(b):

Based on the hinted triangle, we construct the following matrix multiplication as:

$$\begin{bmatrix}-1 & -1 & -1 & 2.5\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

The addition 1 in the vector is the bias. Therefore, only when a data point is inside the triangle, will the product be <0<0, and hence categorized as class 00. Otherwise, it will be classified as class 1. Hence, the accuracy is 100%.

(c):

No, it can't because the current classes cannot be linearly separately. If the activation is linear, then it's an affine transformation from 2D to 3D spaces, and the classification problem will still be non-linearly in the 3D space, which cannot be solved by a step function.