Machine Learning Homework 4

Hairui Yin

1.

(a) Consider two different points x_1, x_2 in the hyperplane $g(x) = 0$, so

$$\begin{cases} w^T x_1 + w_0 = 0 \\ w^T x_2 + w_0 = 0 \end{cases}$$

Subtract two equation, then $w^T(x_1 - x_2) = 0$. It means w^T is the normal vector of that hyperplane.

The distance from a point x_a to the hyperplane means finding a point x^* in the hyperplane ($g(X^*) = 0$), where $x^* - x_a$ is parallel to the normal vector of the hyperplane. So,

$$x^* - x_a = tw, \quad t \in \mathcal{R}$$

This equation can also be written as

$$x^* = x_a + tw, \quad t \in \mathcal{R}$$

Put x^* in to the constraint $g(x) = 0$

$$w^T(x_a + tw) + w_0 = 0$$

After simplification,

$$t = -\frac{w^T x_0 + b}{||w||^2}$$

Therefore, minimize $||x - x_a||^2$ subject to the constraint $g(x) = 0$,

$$\begin{aligned} \text{Distance} &= \sqrt{\min(||x - x_a||^2)} \\ &= \min\left(||x - x_a||\right) \\ &= ||x^* - x_a|| \\ &= ||tw|| \\ &= |t| \cdot ||w|| \\ &= \frac{w^T x_0 + b}{||w||^2} \cdot ||w|| \\ &= \frac{w^T x_0 + b}{||w||} \\ &= \frac{g(x_a)}{||w||} \end{aligned}$$

(b) Since x_p is a projection of x_a onto the hyperplane, $x_p - x_a$ is parallel to the normal vector of the hyperplane. With the inference in question (a), we know that

$$x_p - x_a = tw, \quad t \in \mathcal{R}$$

and

$$t = -\frac{w^T x_0 + b}{||w||^2}$$

Therefore, put t into the above equation,

$$\begin{aligned} x_p &= x_a - \frac{w^T x_0 + b}{||w||^2} w \\ &= x_a - \frac{g(x_a)}{||w||^2} w \end{aligned}$$

2.

(a) After standardizing data dividing by respective standard deviation, we calculate the mean of two classes with specific features

$$m_i = \frac{1}{n_i} \sum_{x \in D_i} x_i$$

Then the within-class scatter matrix is

$$S_w = S_1 + S_2, \quad \text{and} \quad S_i = \sum_{x \in D_i} (x - m_i)(x - m_i)^T$$

Finally, we can calculate w using

$$w = S_w^{-1}(m_1 - m_2)$$

The results of the intermediate steps are shown below:

```
mean of class 1 is  [1.73862886 4.04288329 5.0267895  3.42230893]
mean of class 2 is  [2.61151513 4.87052768 6.58154804 4.69864155]
within-class scatter matrix is
[[1416.87800726  405.5155054    86.38073565   12.62613915]
 [ 405.5155054  1425.3721456    38.40476152  -53.86274044]
 [  86.38073565   38.40476152 1234.11747815  -98.94739238]
 [  12.62613915  -53.86274044  -98.94739238 1321.1452261 ]]
The value of w is
[-0.00039097 -0.00047507 -0.00130421 -0.00107939]
```

(b) The formula of Gini Impurity and Information Gain are

$$\text{Gini Impurity} = 1 - \sum_{j=1}^{c} p_j^2$$

$$\text{information Gain}(S) = \text{Extropy}(S) - \sum \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

The results are shown below:

```
The Gini Impurity of firstly splitting by "hypertension" is  0.14295615100629247
The Gini Impurity of firstly splitting by "heart_disease" is  0.14601404503959087
The information gain of firstly splitting by "hypertension" is  0.014297005653686627
The information gain of firstly splitting by "heart_disease" is  0.003982146278906959
```