9^{th} Homework

Thomas Saltos

Exercise 1

The node impurity is defined as:

$$\mathrm{I}(t) = -\sum_{j=1}^{M} P(\omega_j|t) log_2 \mathrm{P}(\omega_j|t), \qquad P(\omega_j|t) \approx \frac{N_t^j}{N_t}$$

The decrease in node impurity is defined as follows:

$$\Delta\mathrm{I(t)} = \mathrm{I(t)} - \frac{\mathrm{N_{t,Y}}}{\mathrm{N_t}}\mathrm{I(t_Y)} - \frac{\mathrm{N_{t,N}}}{\mathrm{N_t}}\mathrm{I(t_N)}$$

In our case the node impurity is:

- For t:
  It is $P(\omega_1) = \frac{2}{5}$ and $P(\omega_2) = \frac{3}{5}$.
  The entropy is $I(t) = -P(\omega_1|t)log_2 P(\omega_1|t) - P(\omega_2|t)log_2 P(\omega_2|t) = 0.9710$

- For t_Y:

  $$P(\omega_1|t_Y) = \frac{2}{3}, \qquad P(\omega_2|t_Y) = \frac{1}{3}$$

  And

  $$I(t_Y) = -\frac{2}{3}log_2\frac{2}{3} - \frac{1}{3}log_2\frac{1}{3} = 0.918$$

- For t_N:

  $$P(\omega_1|t_N) = 0, \qquad P(\omega_2|t_N) = 1$$

  And

  $$I(t_N) = -0log_2 0 - 1log_2 1 = 0$$

The node impurity decrease:

$$\Delta\mathrm{I(t)} = \mathrm{I(t)} - \frac{\mathrm{N_{t,Y}}}{\mathrm{N_t}}\mathrm{I(t_Y)} - \frac{\mathrm{N_{t,N}}}{\mathrm{N_t}}\mathrm{I(t_N)} = 0.971 - \frac{3}{5}0.918 - \frac{2}{5}0 = 0.42$$

Exercise 2

a) The Lagrangian function is:

$$L(\theta, \theta_0, \lambda) = \frac{1}{2}\theta^T\theta - \sum_{i=1}^{N} \lambda_i [y_i(\theta^T x_i + \theta_0) - 1]$$

Using the Karush-Kuhn-Tacker conditions:

$$\frac{\partial L}{\partial \theta} = 0$$
$$\frac{\partial L}{\partial \theta_0} = 0$$

From the first equation we have:

$$\frac{\partial L}{\partial \theta} = 0 \Rightarrow \theta - \sum_{i=1}^{N} \lambda_i y_i x_i = 0 \Rightarrow \theta = \sum_{i=1}^{N} \lambda_i y_i x_i$$

And

$$\frac{\partial L}{\partial \theta_0} = 0 \Rightarrow \sum_{i=1}^{N} \lambda_i y_i = 0$$

b) We replace to the Lagrangian function:

$$L(\theta, \theta_0, \lambda) = \frac{1}{2}\left(\sum_{i=1}^{N} \lambda_i y_i x_i\right)^T \left(\sum_{i=1}^{N} \lambda_i y_i x_i\right) - \sum_{i=1}^{N} \lambda_i \left[y_i \left(\left(\sum_{i=1}^{N} \lambda_i y_i x_i^T \right) x_i + \theta_0 \right) - 1 \right]$$

Using the equation derived by KKT conditions we have that:

$$L(\lambda_1, \lambda_2, \dots, \lambda_N) = \sum_{i=1}^{N} \lambda_i y_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j x_i^T x_j$$

c) The Wolfe dual representation:

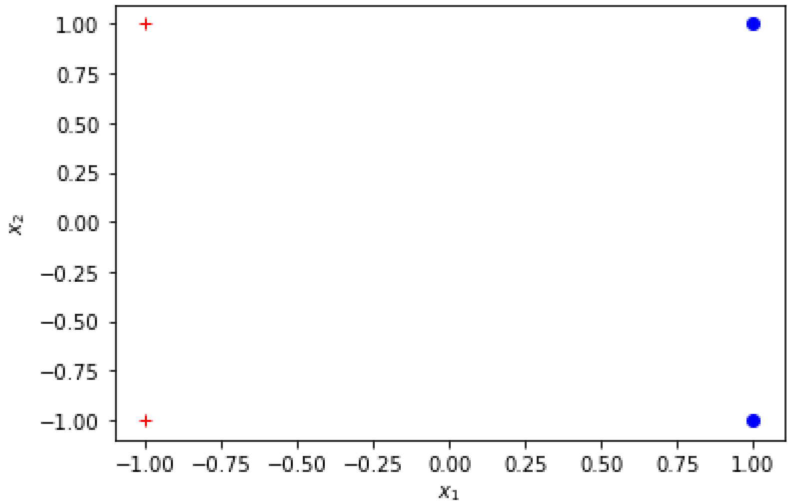
$$max_{\lambda \geq 0} \left(\sum_{i=1}^{N} \lambda_i y_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j x_i^T x_j \right)$$

Subject to $\sum_{i=1}^{N} \lambda_i y_i = 0$.

Exercise 3

a)

```
import matplotlib.pyplot as plt
import numpy as np
X1 = np.array([[-1,1],[-1,-1]])
X2 = np.array([[1,-1],[1,1]])
plt.plot(X1[:,0],X1[:,1],'r+')
plt.plot(X2[:,0],X2[:,1],'bo')
plt.xlabel('$x_1$')
plt.ylabel('$x_2$')
plt.show()
```



It can be noticed that the two classes are linearly separable, and the optimal hyperplane is x_1=0.

b)

i) Lagrange multipliers

In this case we have:

$$N = 4,\ y_1 = y_2 = 1,\ y_3 = y_4 = -1$$

$$y_1 y_1 x_1^T x_1 = 2$$
$$y_1 y_2 x_1^T x_2 = 0$$
$$y_1 y_3 x_1^T x_3 = 2$$
$$y_1 y_4 x_1^T x_0 = 0$$

$$y_2 y_2 x_2^T x_2 = 2$$
$$y_2 y_3 x_2^T x_3 = 0$$
$$y_2 y_4 x_2^T x_4 = 2$$

$$y_3 y_3 x_3^T x_3 = 2$$
$$y_3 y_4 x_3^T x_4 = 0$$

$$y_4 y_4 x_4^T x_4 = 2$$

The Langrangian function is:

$$L = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 - \lambda_1^2 - \lambda_2^2 - \lambda_3^2 - \lambda_4^2 - 2\lambda_1\lambda_3 - 2\lambda_2\lambda_4$$

Taking the gradient with respect to $\lambda_1, \lambda_2, \lambda_3, \lambda_4$

$$\frac{\partial L}{\partial \lambda_1} = 1 - 2\lambda_1 - 2\lambda_3$$
$$\frac{\partial L}{\partial \lambda_2} = 1 - 2\lambda_2 - 2\lambda_4$$
$$\frac{\partial L}{\partial \lambda_3} = 1 - 2\lambda_1 - 2\lambda_3$$

$$\frac{\partial L}{\partial \lambda_4} = 1 - 2\lambda_2 - 2\lambda_4$$

Setting the gradient to 0, we get two equations:

$$\lambda_1 + \lambda_3 = \frac{1}{2}$$
$$\lambda_2 + \lambda_4 = \frac{1}{2}$$

In addition:

$$\sum_{i=1}^{N} \lambda_i y_i = 0 \Rightarrow \lambda_1 + \lambda_2 = \lambda_3 + \lambda_4 \Rightarrow \lambda_1 = \lambda_4 and \lambda_2 = \lambda_3 0 \leq \lambda_i \leq \frac{1}{2}, i = 1,2,3,4$$

ii) The line:

Let $w \in \left[0, \frac{1}{2}\right] \Rightarrow \lambda_1 = \lambda_4 = w,\ \lambda_2 - \lambda_3 = \frac{1}{2} - w$

$$\theta = w[-1,1]^T + \left(\frac{1}{2} - w\right)[-1,-1]^T - \left(\frac{1}{2} - w\right)[1,-1]^T - w[1,1]^T = [-1,0]^T$$

For θ_0:

$$\lambda_i[y_i(\theta^T x_i + \theta_0) - 1] = 0 \Rightarrow [y_i(\theta^T x_i + \theta_0) - 1] = 0$$

More specifically we have:

$$1(-\theta_1 + \theta_2 + \theta_0) - 1 = 0$$
$$1(-\theta_1 - \theta_2 + \theta_0) - 1 = 0$$
$$-1(\theta_1 - \theta_2 + \theta_0) - 1 = 0$$
$$-1(\theta_1 + \theta_2 + \theta_0) - 1 = 0$$

Substituting the values of θ in the above equations we get $\theta_0 = 0$

c) It can be noticed that the solution θ is unique but the resulting Lagrange multipliers (λ_i) are not unique.