

Machine Learning

Sommersemester2020

Exercise 6

Ciheng Zhang (3473188) zch3183505@gmail.com
Gang Yu(3488292) HansVonCq@gmail.com
Huipanjun Tian (3471607) Thpjpyl5111217@gmail.com

June 11, 2020

1 Concept

1. When the dataset is linear separability, Points can be separated by a line. But there is also possible, that separated with a very narrow margin.
2. So we use the slack variable to solve the problem, that separated with narrow margin. we also use slack variable solve some dataset, that most Points can be linear separated, but there are also one or 2 Points, that make a linear separability problem to linear indivisible. slack variable make more fault tolerance to the SVM. we use slack variable to figure out the outlier, and let those outlier no contribute to our loss, that can let margin bigger.
3. Besides the linear separability problem. there are also linear indivisible problem. We use kernel function to solve it. kernel function increase the dimension to make the dataset linear separability in higher dimension.

2 Perceptron

2.1

$$y = \text{sign}(w^T x)$$

2.2

$$w = [1 \quad -1 \quad 0.5], wx_1 = 0.5, y_1 = -1$$

$$w := w - \alpha x_i \text{sign}(f(\hat{x}_i)) = [1 \quad -1 \quad -0.1]$$

$$wx_2 = -1.1, y_2 = 1$$

$$wx_2 * y_2 = -0.5 < 0$$

$$w := [1 \quad -1 \quad 0.5] - 0.6[0 \quad 1 \quad 1] * (-1) = [1 \quad -1.6 \quad 1.1]$$

$$wx_3 = -0.1, y_3 = 1$$

$$wx_3 * y_3 < 0$$

$$w := [1 \quad -1.6 \quad 1.1] - 0.6 * [1 \quad 0 \quad 1] * (-1) = [1.6 \quad -1.6 \quad 1.7]$$

$$wx_4 = 1.7, y_4 = 1$$

$$wx_4 * y_4 > 0$$

$$wx_1 = 1.7, y_1 = -1$$

$$wx_1 * y_1 < 0$$

$$w := [1.6 \quad -1.6 \quad 1.7] - 0.6[0 \quad 0 \quad 1] * (1) = [1.6 \quad -1.6 \quad 1.1]$$

$$wx_1 = 1.1, y_1 = -1$$

$$wx_1 * y_1 < 0$$

$$w := [1.6 \quad -1.6 \quad 1.1] - 0.6[0 \quad 0 \quad 1] * (1) = [1.6 \quad -1.6 \quad 0.5]$$

$$w := [1.6 \quad -1.6 \quad 0.5] - 0.6[0 \quad 0 \quad 1] * (1) = [1.6 \quad -1.6 \quad -0.1]$$

$$wx_2 = -1.7$$

$$w := [1.6 \quad -1.1 \quad 0.5]$$

$$wx_1 = 0.5$$

$$w := [1.6 \quad -1.1 \quad -0.1]$$

$$wx_2 = -1.2$$

$$w := [1.6 \quad -0.5 \quad 0.5]$$

$$wx_1 = 0.5$$

$$w := [1.6 \quad -0.5 \quad -0.1]$$

$$wx_2 = -0.7$$

$$w := [1.6 \quad 0.1 \quad 0.5]$$

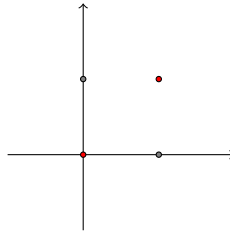
$$wx_1 = 0.5$$

$$w := [1.6 \quad 0.1 \quad -0.1]$$

Then every Points can make $\hat{f}(x_i)y_i < 0$

$$\hat{f}(x_i) = \text{sign}(1.6x_1 + 0.1x_2 - 0.1)$$

2.3



acording to the picture, we can konw the XOR problem cannot be represented by a (linear) Perceptron.

3 Polynomial Kernel

$$\begin{aligned}\Phi(x_i)^T \Phi(x_j) &= [x_{i1}^2 \quad \sqrt{2}x_{i1}x_{i2} \quad x_{i2}^2] \begin{bmatrix} x_{j1}^2 \\ \sqrt{2}x_{j1}x_{j2} \\ x_{j2}^2 \end{bmatrix} \\ &= (x_{i1}x_{j1})^2 + 2x_{i1}x_{i2}x_{j1}x_{j2} + (x_{i2}x_{j2})^2 \\ &= (x_{i1}x_{j1} + x_{i2}x_{j2})^2\end{aligned}$$

So we don't need to calculate the scalar product, and we can use $(x_{i1}x_{j1} + x_{i2}x_{j2})^2$ to instead of it.

4 Gaussian kernel

form slide 69 we got:

$$\begin{aligned}k(x, x') &= e^{-\frac{\|x-x'\|^2}{2\sigma^2}} \\ &= e^{\frac{-1}{2\sigma}} e^{-x^2} e^{-x'^2} e^{2xx'}\end{aligned}$$

Then we use Tayler Expansion:

$$\begin{aligned}&= e^{\frac{-1}{2\sigma}} e^{-x^2} e^{-x'^2} \sum_{i=0}^{\infty} \frac{(2xx')^i}{i!} \\ &= e^{\frac{-1}{2\sigma}} \sum_{i=0}^{\infty} e^{-x^2} e^{-x'^2} \frac{(2xx')^i}{i!} \\ &= e^{\frac{-1}{2\sigma}} \sum_{i=0}^{\infty} e^{-x^2} e^{-x'^2} \sqrt{\frac{2^i}{i!}} \sqrt{\frac{2^i}{i!}} x^i x'^i \\ &= \Phi(x)^T \Phi(x') \\ \Phi(x) &= e^{-x^2} * (1 + \sqrt{\frac{2}{1!}}x + \sqrt{\frac{2^2}{2!}}x^2 + \dots)\end{aligned}$$

So its infnty dimension.