Machine Learning Sommersemester2020 Exercise 4

Ciheng Zhang (3472321) zch3183505@gmail.com
Gang Yu(3488292) HansVonCq@gmail.com
Huipanjun Tian (3471607) Thpjpyl5111217@gmail.com

May 24, 2020

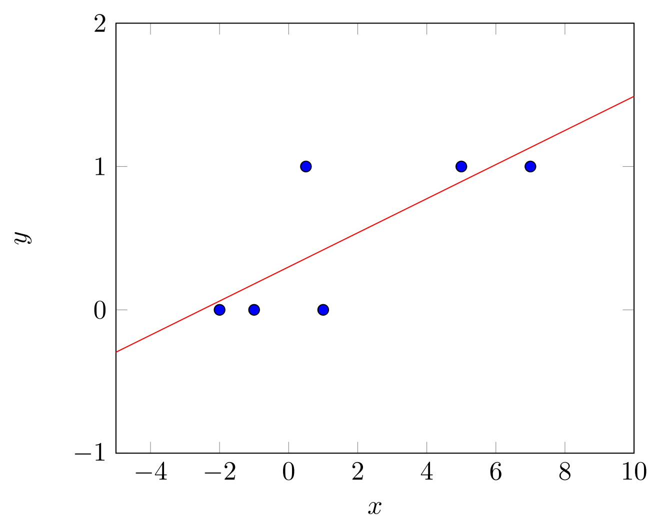
Ciheng Zhang(3472321)
Gang Yu(3488292)
Huibanjun Tian(3471607)

1 Classification with Linear Regression

$$x = [-2.0, -1.0, 0.5, 0.6, 5.0, 7.0]^T$$
$$y = [0, 0, 1, 0, 1, 1]^T$$
$$y_i = \beta_0 + \beta_1 x_i + \epsilon$$
$$\epsilon = y_i - \beta_0 - \beta_1 x_i$$
$$\beta = (\beta_0, \beta_1)^T$$
$$\overline{x}_i = (1, x)^T$$
$$X = \begin{bmatrix} 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \\ -2 & -1 & 0.5 & 0.6 & 5.0 & 7.0 \end{bmatrix}^T$$
$$L^{ls}(\beta) = ||\epsilon(\beta)||^2 = ||y - X\beta||^2$$

we should minimal the loss function$L^{ls}(\beta)$:

$$\frac{\partial L^{ls}(\beta)}{\partial \beta} = -2(y - X\beta)^T X = 0_d^T$$
$$\hat{\beta}^{ls} = (X^T X)^{-1} X^T y = [0.2997, 0.1190]$$
$$\hat{y} = 0.1190x + 0.2997$$



So the point $(-2, 0), (-1, 0), (0.6, 0), (7, 0)$ is one class,the other points is another class. We decide a point belong to witch class, due to it upper or under the line. Because the anwser of the linear regression is continue value not the proability and its really sensitive to the disturbute of the data, So the linear regression is not suitable for classification.

2 Log-likelihood gradient and Hessian

$$L(\beta) = \sum_{i=1}^{N} [y_i logp(x_i) + (1 - y_i) log[1 - p(x_i)]]$$

$$\frac{\partial}{\partial \beta} L(\beta) = \sum_{i=1}^{N} [\frac{\partial}{\partial \beta} y_i logp(x_i) + \frac{\partial}{\partial \beta}(1 - y_i)(log(1 - p(x_i)))]$$

$$\frac{\partial}{\partial \beta} y_i logp(x_i) = (\frac{\partial}{\partial \beta} y_i) logp(x_i) + y_i(\frac{\partial}{\partial \beta} logp(x_i))$$

$$p(x) = \sigma(f(x))$$

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

$$\frac{\partial}{\partial z} p(z) = p(z)(1 - p(z)) f'(z)$$

$$\frac{\partial}{\partial z} logp(z) = \frac{p(z)(1 - p(z)) f'(z)}{p(z)} = (1 - p(z)) f'(z)$$

$$\frac{\partial}{\partial \beta} y_i logp(x_i) = 0 + y_i(1 - p(x_i))\phi(x_i)^T$$

$$\frac{\partial}{\partial \beta}(1 - y_i) log[1 - p(x_i)] = (1 - y_i)\frac{\partial}{\partial \beta} log[1 - p(x_i)]$$

$$= -p(z) f'(z)$$

$$\frac{\partial}{\partial \beta} L(\beta) = \sum_{i=1}^{N} y_i(1 - p(x_i))\phi(x_i)^T - (1 - y_i) p(x_i)\phi(x_i)^T$$

$$= \sum_{i=1}^{N} x_i(y_i - p(x_i))$$

$$\frac{\partial^2}{\partial \beta^2} L(\beta) = \sum_{i=1}^{N} x_i \frac{\partial}{\partial \beta}(y_i - p(x_i)) + (y_i - p(x_i))\frac{\partial x_i}{\partial \beta}$$

$$\frac{\partial x_i}{\partial \beta} = \frac{\partial y_i}{\partial \beta} = 0$$

$$\frac{\partial}{\partial \beta} p(x_i) = p(x_i)(1 - p(x_i)) f'(x_i) = p(x_i)(1 - p(x_i)) x_i$$

$$\frac{\partial^2}{\partial \beta^2} L(\beta) = \sum_{i=1}^{N} x_i^2 p(x_i)(p(x_i) - 1)$$

Ciheng Zhang(3472321)
Gang Yu(3488292)
Huibanjun Tian(3471607)

3 Discriminative Function in Logistic Regression

$$f(x, y) = \Phi(x, y)^T \beta$$

$$\Phi(x, y) = \Phi(x)[y = 1]$$

In the binary case $f(x, 0)$ and $f(x, 1)$. Then $f'(x, 0) = 0$.

$$f(x, 1) = \Phi(x)^T \beta$$

$$p(y = 1|x) = \frac{e^{f(x,1)}}{e^{f(x,0)} + e^{f(x,1)}} = \frac{e^{f(x,1)}}{1 + e^{f(x,1)}} = \sigma(f(x, 1))$$

According to the above calculation, we can get, we can assume $f(x, 0) = 0$ without loss generality.