Modern Multivariate Statistical Techniques

Jinhong Du,15338039

January 7, 2019

Content

1 Ex 10.6

Consider four points, (X_1, X_2), at the corners of the unit square: $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$. Suppose that $(0,0)$ and $(1,1)$ are in class 1, whereas $(0,1)$ and $(1,0)$ are in class 2. The XOR problem is to construct a network that classifies the four points correctly. By setting $Y = 1$ to points in class 1 and $Y = 0$ to points in class 2 (or vice versa), show algebraically that a straight line cannot separate the two classes of points and, hence, that a perceptron with no hidden nodes is not an appropriate network for this problem.

A perceptron with weight w and bias b can be represented by

$$f(x) = \begin{cases} 1 & , w^\top x + b \leq 0 \\ 2 & , w^\top x + b > \end{cases}$$

Suppose that these four points can be classified correctly, then we have

$$\begin{bmatrix} 0 & 0 \end{bmatrix} w + b \leq 0 \tag{1}$$
$$\begin{bmatrix} 1 & 1 \end{bmatrix} w + b \leq 0 \tag{2}$$
$$\begin{bmatrix} 0 & 1 \end{bmatrix} w + b > 0 \tag{3}$$
$$\begin{bmatrix} 1 & 0 \end{bmatrix} w + b > 0 \tag{4}$$

By (1)+(2) and (3)+(4), we have

$$\begin{bmatrix} 1 & 1 \end{bmatrix} w + 2b \leq 0$$
$$\begin{bmatrix} 1 & 1 \end{bmatrix} w + 2b > 0$$

which contradicts each other. Therefore, a perceptron with no hidden nodes is not an appropriate network for this problem.

2 Ex 10.7

Consider a fully connected network with two input nodes (X_1, X_2), two hidden nodes (Z_1, Z_2), and a single output node (Y). Let $\beta_{11} = \beta_{12} = 1$ be the connection weights from X_1 to Z_1 and Z_2, respectively; let $\beta_{01} = 1.5$ be the bias at hidden node 1; let $\beta_{21} = \beta_{22} = 1$ be the connection weights from X_2 to Z_1 and Z_2, respectively; and let $\beta_{02} = 0.5$ be the bias at hidden node 2. Next, let $\alpha_1 = -2$ and $\alpha_2 = 1$ be the connection weights from Z_1 to Y and from Z_2 to Y, respectively, with bias $\alpha_0 = 0.5$. Draw the network graph. Find the linear boundaries as defined by the two hidden nodes; in the unit square, draw the boundaries and identify which class, 0 or 1, corresponds to each region of the unit square. Show that this network solves the XOR problem. Find another solution to this problem using different weights and biases.

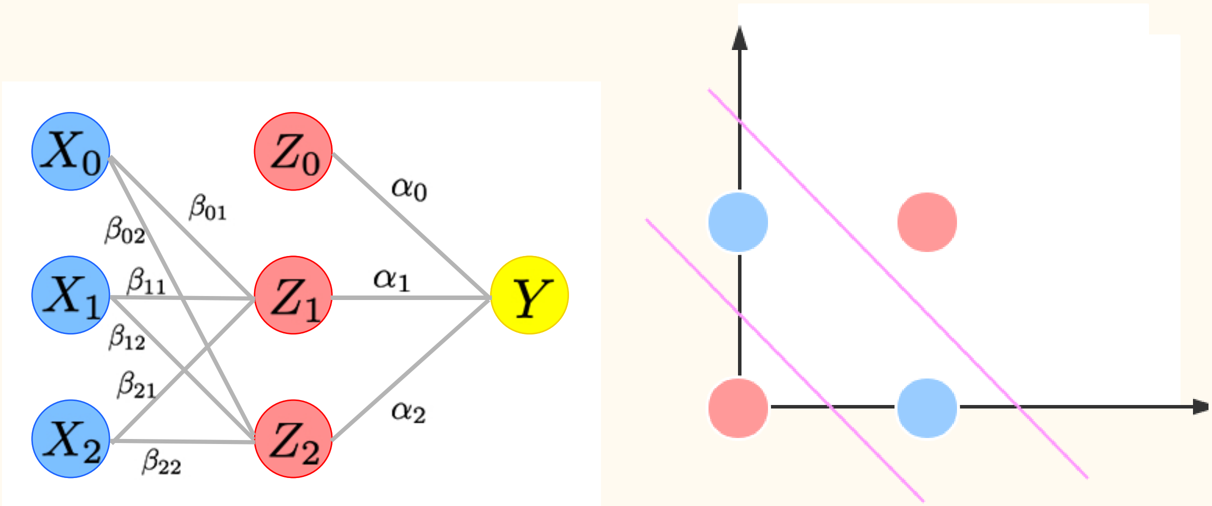

图 1: Structure of Nerual Network

$$
\begin{aligned}
Z_1 &= \beta_{01}X_0 + \beta_{11}X_1 + \beta_{21}X_2 \\
&= 1.5 + X_1 + X_2 \\
Z_2 &= \beta_{02}X_0 + \beta_{12}X_1 + \beta_{22}X_2 \\
&= 0.5 + X_1 + X_2 \\
Y &= \alpha_0 Z_0 + \alpha_1 Z_1 + \alpha_2 Z_2 \\
&= 0.5 - 2(1.5 + X_1 + X_2) + (0.5 + X_1 + X_2) \\
&= -2 - X_1 - X_2
\end{aligned}
$$

When $\mathbf{X} = (0,0)$, $Y = -2$.
When $\mathbf{X} = (1,1)$, $Y = -4$.
When $\mathbf{X} = (0,1)$, $Y = -3$.
When $\mathbf{X} = (1,0)$, $Y = -3$.

3 Ex 11.11

Show that the functional $F_D(\boldsymbol{\alpha})$ in (11.40) is concave; i.e., show that, for $\theta \in (0,1)$ and $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^n$,

$$F_D(\theta\boldsymbol{\alpha} + (1-\theta)\boldsymbol{\beta}) \geq \theta F_D(\boldsymbol{\alpha}) + (1-\theta)F_D(\boldsymbol{\beta})$$

where

$$F_D(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^\top \mathbf{x}_j)$$

$\because$

$$F_D(\boldsymbol{\alpha}) = \mathbf{1}^\top \boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^\top \mathbf{H}\boldsymbol{\alpha}$$

where

$$\mathbf{H} = \begin{bmatrix} \langle y_1\mathbf{x}_1, y_1\mathbf{x}_1\rangle & \cdots & \langle y_1\mathbf{x}_1, y_n\mathbf{x}_n\rangle \\ \vdots & \ddots & \vdots \\ \langle y_n\mathbf{x}_n, y_1\mathbf{x}_1\rangle & \cdots & \langle y_n\mathbf{x}_n, y_n\mathbf{x}_n\rangle \end{bmatrix}$$

$\therefore$

$$\begin{aligned} F_D(\theta\boldsymbol{\alpha} + (1-\theta)\boldsymbol{\beta}) &= \mathbf{1}^\top[\theta\boldsymbol{\alpha} + (1-\theta)\boldsymbol{\beta}] - \frac{1}{2}[\theta\boldsymbol{\alpha} + (1-\theta)\boldsymbol{\beta}]^\top \mathbf{H}[\theta\boldsymbol{\alpha} + (1-\theta)\boldsymbol{\beta}] \\ &= \mathbf{1}^\top(\theta\boldsymbol{\alpha}) - \frac{1}{2}(\theta\boldsymbol{\alpha})^\top \mathbf{H}(\theta\boldsymbol{\alpha}) \\ &\quad + \mathbf{1}^\top[(1-\theta)\boldsymbol{\beta}] - \frac{1}{2}[(1-\theta)\boldsymbol{\beta}]^\top \mathbf{H}[(1-\theta)\boldsymbol{\beta}] \\ &\quad - \frac{1}{2}(\theta\boldsymbol{\alpha})^\top \mathbf{H}[(1-\theta)\boldsymbol{\beta}] - \frac{1}{2}[(1-\theta)\boldsymbol{\beta}]^\top \mathbf{H}(\theta\boldsymbol{\alpha}) \\ &= \theta\mathbf{1}^\top\boldsymbol{\alpha} - \theta\frac{1}{2}\boldsymbol{\alpha}^\top\mathbf{H}\boldsymbol{\alpha} \\ &\quad + (1-\theta)\mathbf{1}^\top\boldsymbol{\beta} - (1-\theta)\frac{1}{2}\boldsymbol{\beta}^\top\mathbf{H}\boldsymbol{\beta} \\ &\quad + \frac{1}{2}\theta(1-\theta)(\boldsymbol{\alpha}-\boldsymbol{\beta})^\top\mathbf{H}(\boldsymbol{\alpha}-\boldsymbol{\beta}) \\ &= \theta F_D(\boldsymbol{\alpha}) + (1-\theta)F_D(\boldsymbol{\beta}) + \frac{1}{2}\theta(1-\theta)(\boldsymbol{\alpha}-\boldsymbol{\beta})^\top\mathbf{H}(\boldsymbol{\alpha}-\boldsymbol{\beta}) \end{aligned}$$

Now we need to prove that H is semi-positive definite. Since $\forall\ \mathbf{z} \in \mathbb{R}^n$

$$\begin{aligned} \mathbf{z}^\top\mathbf{H}\mathbf{z} &= \sum_{i=1}^n \sum_{j=1}^n z_i z_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j\rangle \\ &= \sum_{i=1}^n z_i y_i \langle \mathbf{x}_i, \sum_{j=1}^n z_j y_j \mathbf{x}_j\rangle \\ &= \langle \sum_{i=1}^n z_i y_i \mathbf{x}_i, \sum_{j=1}^n z_j y_j \mathbf{x}_j\rangle \\ &\geq 0 \end{aligned}$$

therefore **H** is semi-positive definite. So,

$$(\boldsymbol{\alpha}-\boldsymbol{\beta})^\top\mathbf{H}(\boldsymbol{\alpha}-\boldsymbol{\beta}) \geq 0$$

and therefore,

$$F_D(\theta\boldsymbol{\alpha} + (1-\theta)\boldsymbol{\beta}) \geq \theta F_D(\boldsymbol{\alpha}) + (1-\theta)F_D(\boldsymbol{\beta})$$