

EMS702P Statistical Thinking and Applied Machine Learning

Week 8 – PBL

Yunpeng Zhu

Q1: (25 min) Classification

Consider we have 4 sets of observed data

$$(x_1, x_2) = [(0.5, 1); (1, 0.5); (2, 2); (2, 1)]$$

$$y = \{1, 1, 0, 0\}$$

Evaluate the classifier as a linear polynomial function:

$$y = 2 + a_1 x_1 + a_2 x_2$$

1. Formulate the classifier model into a matrix form and show the Least Squares (LS) representation of a_1 and a_2 ;
2. Estimate the values of a_1, a_2 from the observed data by using the LS method.
3. Evaluate the classification boundary.

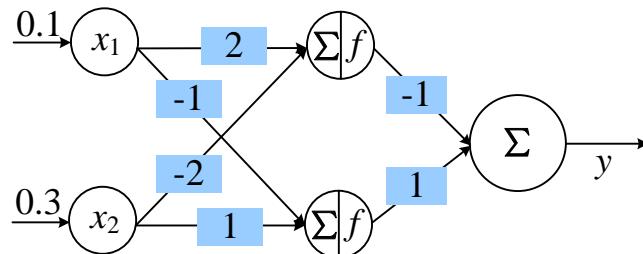
Q2: (10 min)

Find the distance from the point $(x_1, x_2) = (1, 1)$ to the line $x_1 + x_2 - 1 = 0$

$$x_1 + x_2 - 1 = 0 \Rightarrow \bar{\mathbf{x}}^T \bar{\mathbf{w}} + w_0 = 0: [x_1 \quad x_2] \times \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 1 = 0$$

$$d = \frac{w_0 + \mathbf{x}_0^T \bar{\mathbf{w}}}{\|\bar{\mathbf{w}}\|_2} = \frac{-1 + [1 \quad 1] \times \begin{bmatrix} 1 \\ 1 \end{bmatrix}}{\sqrt{1^2 + 1^2}} = \frac{\sqrt{2}}{2}$$

Q3: (10 min)



$$f(z) = \max(0, z) . \text{Calculate } y$$