L- class classification problem

$(X_i, y_i),\ i = 1, 2, \cdots n,\ y_i \in \{1, 2, \cdots L\}$

Goal: predict the response for a new X_0

Step1: Find the K-nearest-neighbors.:

compute the distance $d_i = \|X_0 - X_i\|_2$

$= \sqrt{\sum_{\ell=1}^{p} (X_{0\ell} - X_{i\ell})^2}$, for p-dimensional X

(Euclidean Distance)

Find set N_0 for the K smallest d_i values

$K = 5 \Rightarrow N_0 = \{1, 5, 8, 23, 56\}$

Step 2: (For regression: $\hat{y}_0 = \frac{1}{5}(y_1 + y_5 + y_8 + y_{23} + y_{56})$)

$$\boxed{\hat{P}(y_0 = \ell \mid X = x_0) = \frac{1}{K} \boxed{\sum_{i \in N_0} 1\{y_i = \ell\}}}$$

$y_1 = 2,\ y_5 = 3,\ y_8 = 2,\ y_{23} = 2,\ y_{56} = \underline{1}$

$L = 3$

$\hat{p}(y_0 = 1 \mid X = x_0) = \frac{1}{5}$

$\hat{p}(y_0 = 2 \mid X = x_0) = \frac{3}{5} \Rightarrow \hat{y}_0 = 2$

$\hat{p}(y_0 = 3 \mid X = x_0) = \frac{1}{5}$

$$\hat{y}_0 = \arg\max_{\ell} \hat{p}(y_0 = \ell \mid X = x_0)$$

$$\boxed{\hat{y}_0 = \underset{\ell}{\arg\max}\, \hat{P}(y_0 = \ell \mid X = x_0)}$$

LR:

$$P(y_0 = 1 \mid X = x_0) = \frac{e^{\beta_0 + \beta_1 x_0}}{1 + e^{\beta_0 + \beta_1 x_0}}$$

$$\hat{p}(y_0 = 1 \mid X = x_0) > 0.5 \Rightarrow \hat{y}_0 = 1$$

$$\Longleftrightarrow \hat{p}(y_0 = 1 \mid X = x_0) > \hat{p}(y_0 = 0 \mid X = x_0)$$

$$\Longleftrightarrow 1 = \underset{\ell}{\arg\max}\, \hat{p}(y_0 = \ell \mid X = x_0)$$

LDA/QDA: same thing!

$X_1 = (100, 1)$

$X_2 = (80, 5)$

$X_3 = (70, 9)$

$X_4 = (80, 1)$

$$\|X_1 - X_2\| = \sqrt{(100-80)^2 + (1-5)^2} = \sqrt{400 + 16}$$

$\downarrow$ standardize the variables

$\tilde{X}_1 = (10, 1)$

$\tilde{X}_2 = (8, 5)$

$\tilde{X}_3 = (7, 9)$

$\tilde{X}_4 = (8, 1)$

$\rightarrow$ toy example