13

Linear and Logistic Regression

§13.1 Linear and Logistic Regression

(Regression Function) $r(x) := \mathbb{E}(Y \mid X=x) = \int y \cdot f(y \mid x) dy$.

Goal: estimate $r(x)$ from given observation
$(Y_1, X_1), \cdots, (Y_n, X_n) \sim F_{X,Y}$.

Prediction: estimate Y for a new observation

§13.2 Simple Linear Regression

When X_i is simple (one-dimensional), and $r(x)$ assumed to be linear: $r(x) = \beta_0 + \beta_1 x$

Suppose $\mathbb{V}(\varepsilon_i \mid X=x) = \sigma^2$ does not depend on x.

$\boxed{\text{Definition}}$ (Simple Linear Regression Model)

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \text{ where } \begin{cases} \mathbb{E}(\varepsilon_i \mid X_i) = 0. \\ \mathbb{V}(\varepsilon_i \mid X_i) = \sigma^2. \end{cases}$$

$\hat{\beta}_0, \hat{\beta}_1$: estimates of β_0, β_1.

$\hat{\varepsilon}_i = Y_i - \hat{Y}_i = Y_i - \hat{r}(X_i) = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)$. (Residuals)

$\boxed{\text{Defininition}}$ (Least Square Estimate)

Values $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize $RSS = \sum_{i=1} \hat{\varepsilon}_i^2$.

$\boxed{\text{Theorem}}$ (Least Square Estimate in Closed Form)

$$\hat{\beta}_1 = \frac{\sum (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\sum (X_i - \overline{X}_n)^2}$$

$$\hat{\beta}_0 = \overline{Y}_n - \hat{\beta}_1 \overline{X}_n.$$

$$\hat{\sigma}^2 = \left(\frac{1}{n-2}\right) \sum \hat{\varepsilon}_i^2$$

(pf) $RSS = \sum \hat{\varepsilon}_i^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = -2 \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = -2 \sum X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0.$$

$$\overline{Y}_n - \hat{\beta}_0 - \hat{\beta}_1 \overline{X}_n = 0.$$

$$\sum X_i Y_i = \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2.$$

$$= (\bar{Y}_n - \bar{X}_n \hat{\beta}_1) \sum X_i + \hat{\beta}_1 \sum X_i^2 .$$

$$\hat{\beta}_1 = \frac{\sum X_i Y_i - \bar{Y}_n \sum X_i}{\sum X_i^2 - \bar{X}_n \sum X_i} = \frac{\sum (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sum (X_i - \bar{X}_n)^2}$$

$$\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{X}_n .$$

$$\hat{\sigma}^2 = ?$$

§13.2 Least Squares and Maximum Likelihood

Assume $\varepsilon_i | X_i \sim N(0, \sigma^2)$.

i.e. $Y_i | X_i \sim N(\mu_i, \sigma^2)$ for $\mu_i = \beta_0 + \beta_1 X_i$.

Likelihood ftn: $\prod_{i=1}^{n} f(X_i, Y_i) = \prod_{i=1}^{n} f_X(X_i) f_{Y|X}(Y_i | X_i)$

$$= \prod f_X(X_i) \prod f_{Y|X}(Y_i|X_i)$$

$$= \mathcal{L}_1 \times \mathcal{L}_2.$$

$\mathcal{L}_2$: conditional likelihood.

$$\mathcal{L}_2 = \mathcal{L}(\beta_0, \beta_1, \sigma) = \prod_{i=1}^{n} f_{Y|X}(Y_i | X_i) \propto \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2} \sum (Y_i - \mu_i)^2\right).$$

$$\ell(\beta_0, \beta_1, \sigma) = -n \log \sigma - \frac{1}{2\sigma^2} \sum \left(Y_i - (\beta_0 + \beta_1 X_i)\right)^2.$$

MLE of (β_0, β_1) maximize $\ell(\beta_0, \beta_1, \sigma)$, and thus minimize RSS.

→ $\boxed{\text{Theorem}}$ Under assumption of normality,

Least square estimator $\equiv$ max. likelihood estimator.

§13.3 Properties of the Least Squares Estimators

$\boxed{\text{Theorem}}$ Let $\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$ denote the least squares estimators.

Then, $\mathbb{E}(\hat{\beta} | X^n) = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$

$$\mathbb{V}(\hat{\beta} | X^n) = \frac{\sigma^2}{n S_x^2} \begin{pmatrix} \frac{1}{n} \sum X_i^2 & -\bar{X}_n \\ -\bar{X}_n & 1 \end{pmatrix}$$

where $S_x^2 = n^{-1} \sum (X_i - \bar{X}_n)^2$.

(pf)

$\boxed{\text{Theorem}}$ Under appropriate conditions,

1. $\hat{\beta}_0 \xrightarrow{P} \beta_0$, $\hat{\beta}_1 \xrightarrow{P} \beta_1$

2. $\dfrac{\hat{\beta}_0 - \beta_0}{\hat{se}(\hat{\beta}_0)} \rightsquigarrow N(0,1)$, $\dfrac{\hat{\beta}_1 - \beta_1}{\hat{se}(\hat{\beta}_1)} \rightsquigarrow N(0,1)$

3. $1-\alpha$ confidence interval for β_0, β_1 are
   $\hat{\beta}_0 \pm z_{\alpha/2} \hat{se}(\hat{\beta}_0)$, $\hat{\beta}_1 \pm z_{\alpha/2} \hat{se}(\hat{\beta}_1)$.

4. The Wald test for $H_0: \beta_1 = 0$, vs $H_1: \beta_1 \neq 0$ is reject H_0 if $|W| > z_{\alpha/2}$ with $W = \hat{\beta}_1 / \hat{se}(\hat{\beta}_1)$.

(pf)

§13.4 Prediction

$\boxed{\text{Theorem}}$ (Prediction Interval)

Let $\hat{\xi}_n^2 = \hat{\sigma}^2 \left(\dfrac{\Sigma (X_i - X_*)^2}{n \Sigma (X_i - \bar{X})^2} + 1 \right)$.

An approximate $1-\alpha$ prediction interval for Y_* is

$$Y_* \pm z_{\alpha/2} \hat{\xi}_n$$

(pf)

$\boxed{\text{Note}}$ $\mathbb{E}(Y_* | X_*) = \beta_0 + \beta_1 X_*$.

$$\mathbb{E}(\hat{Y}_* | X_*) = \mathbb{E}(Y_* | X_*)$$

$$\hat{Y}_* - \mathbb{E}(Y_* | X_*) = (\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1) X_*$$

$$\mathbb{V}(\hat{Y}_* | X_*) = \mathbb{V}(\hat{\beta}_0 + \hat{\beta}_1 X_*)$$

$$= \mathbb{V}(\hat{\beta}_0) + X_*^2 \mathbb{V}(\hat{\beta}_1) + 2\,\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1)$$

Banging away on piano
Sending away the semester

Tute 1.

§13.5 Multiple Regression

Data: $(Y_1, X_1), \cdots, (Y_n, X_n)$,
$X_i = (X_{i1}, \cdots, X_{ik})$.

Linear Regression Model: $Y_i = \sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i$.

$\mathbb{E}(\varepsilon_i | X_{1i}, \cdots, X_{ki}) = 0$.

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \quad X = \begin{pmatrix} X_{11} & \cdots & X_{1k} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{nk} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \quad Y = X\beta + \varepsilon.$$

§13.6 Model Selection