STATS 305B: Methods for Applied Statistics I — Winter 2016/17

Lecture 8: January 27

Lecturer: Jonathan Taylor — *Scribes: Kenneth Tay*

8.1 Population Least Squares

8.1.1 Linear Regression

Under the usual Gaussian model for linear regression, we assume that $Y \mid X \sim \mathcal{N}(X\beta, \sigma^2 I)$.

Recall that when trying to minimize least squares, $\hat{\beta}$ has to satisfy the Normal equations

$$\begin{aligned} X^T(Y - X\hat{\beta}) &= 0, \\ \hat{\beta} &= (X^T X)^{-1} X^T Y. \end{aligned}$$

When the model is not Gaussian, say $(X_i, Y_i) \overset{i.i.d.}{\sim} F$ where F is some distribution, $\hat{\beta}$ is estimating β which satisfies

$$\begin{aligned} \mathbb{E}_F[X^T(Y - X\beta)] &= 0, \\ \beta &= (\mathbb{E}_F[X^T X])^{-1} \mathbb{E}_F[X^T Y]. \end{aligned}$$

Let $\varepsilon = Y - X^T\beta$ (i.e. residuals under the true model). We have

$$\begin{aligned} \hat{\beta} &= (X^T X)^{-1} X^T Y \\ &= (X^T X)^{-1} X^T (X\beta + \varepsilon) \\ &= \beta + (X^T X)^{-1} X^T \varepsilon, \\ \hat{\beta} - \beta &= (X^T X)^{-1} X^T \varepsilon. \end{aligned}$$

Note that $X^T\varepsilon = \displaystyle\sum_{i=1}^n \varepsilon_i X_i$, which (by CLT) is eventually normally distributed like $\mathcal{N}(0, \Sigma_F)$.

For large n,

$$\begin{aligned} X^T X &= \sum_{i=1}^n X_i X_i^T \\ &\approx n\mathbb{E}_F[XX^T] =: Q_F, \end{aligned}$$

Then

$$\begin{aligned} \hat{\beta} - \beta &= (X^T X)^{-1} X^T \varepsilon \\ &\approx (nQ_F)^{-1} X^T \varepsilon \\ &\approx \mathcal{N}\left(0, (nQ_F)^{-1}(n\Sigma_F)(nQ_F)^{-1}\right) \\ &= \mathcal{N}\left(0, \frac{1}{n} Q_F^{-1} \Sigma_F Q_F^{-1}\right) \end{aligned}$$

The variance above has the so-called "sandwich form".

8.1.2 Logistic Regression

In logistic regression, the equivalent to the Normal equations is

$$X^T[Y - \pi_{\hat{\beta}}(X)] = 0.$$

In the general setting where $(X_i, Y_i) \overset{i.i.d.}{\sim} F$ for some distribution F, $\hat{\beta}$ is estimating β which satisfies

$$\mathbb{E}_F[X^T[Y - \pi_\beta(X)]] = 0.$$