Notes on Machine Learning

Keng-Wit Lim*[1]

[1]XXXX Los Angeles, CA, USA

March 15, 2018

Abstract

Notes on Machine Learning

Keywords: XXX

1 Notation

$$\mathbf{y} = \begin{Bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{Bmatrix} \tag{1}$$

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & \dots & X_{1p} \\ 1 & X_{21} & \dots & X_{2p} \\ & \vdots & & \\ 1 & X_{N1} & \dots & X_{Np} \end{bmatrix} \tag{2}$$

*kengwit@gmail.com

$$\boldsymbol{\beta} = \left\{ \begin{array}{c} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{array} \right\} \tag{3}$$

2 Linear Regression

2.1 Assumptions

The following assumptions will allow us to draw inferences about the estimators and linear regression model:

1. A linear regression model assumes that the regression function

$$\begin{aligned} f(\mathbf{X}) &= \mathrm{E}(\mathbf{y}|\mathbf{X}) \\ &= \beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j \\ &= \mathbf{X}\beta \end{aligned} \tag{4}$$

   is linear in the inputs $\mathbf{X}$. This means that (4) is assumed to be the correct model for the mean and that the conditional expectation of $\mathrm{E}(\mathbf{y}|\mathbf{X})$ is linear in $X_1, \ldots, X_p$.

2. The *true* relation between a quantitative response $\mathbf{y}$ on the basis of predictors $\mathbf{X}$ is assumed to take the form

$$\begin{aligned} \mathbf{y} &= \mathrm{E}(\mathbf{y}|\mathbf{X}) + \boldsymbol{\epsilon} \\ &= \mathbf{X}\beta + \boldsymbol{\epsilon} \end{aligned} \tag{5}$$

   where $\boldsymbol{\epsilon}$ is the error or residual vector, and it is assumed that $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\sigma})$ i.e. vector of zero mean and vector of (unobserved) variance $\boldsymbol{\sigma}$. The assumed relation (5) means that the deviations of $\mathbf{y}$ around its expectation are additive and Gaussian.

2.2 Solution for the Estimators

In linear regression, we assume that there is approximately a linear relation between $\mathbf{y}$ and $\mathbf{X}$:

$$\mathbf{y} \approx \mathbf{X}\beta \tag{6}$$

The least-squares solution for the estimator vector is

$$\hat{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \tag{7}$$

With the estimator vector, the predicted values are:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \tag{8}$$

2.3 Properties of the Estimators

Now, we derive the mean and variance for the estimator. The mean is

$$\mathrm{E}(\hat{\beta}) = \mathrm{E}\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}\right] \tag{9}$$

The variance is

$$\begin{aligned} \mathrm{Var}(\hat{\beta}) &= \mathrm{E}\left[\left(\hat{\beta} - \mathrm{E}(\hat{\beta})\right)^2\right] \\ &= \mathrm{E}(\hat{\beta}^2) - \mathrm{E}(\hat{\beta})\mathrm{E}(\hat{\beta}) \\ &= \mathrm{E}(\hat{\beta}^2) - \beta^2 \quad \text{since } \mathrm{E}(\hat{\beta}) = \beta \end{aligned} \tag{10}$$