Notes on Introduction to Statistical Learning

Vishal Burman

September 5, 2021

1 Linear Regression

Mathematically we can write the linear relationship as:

$$Y \approx \beta_0 + \beta_1 X \tag{1}$$

Once we know the coefficients from training we can predict using:

$$\hat{y} = \hat{\beta_0} + \hat{\beta_1} x \tag{2}$$

1.1 Estimating the Coefficients

We need to estimate $\hat{\beta_0}$ and $\hat{\beta_1}$ in such a way that it is as close as possible to β_0 and β_1. The most common approach is using the least squares criterion.

ith residual is calculated as:

$$e_i = y_i - \hat{y}_i \tag{3}$$

We define the *residual sum of squares* (RSS) as:

$$RSS = e_1^2 + e_2^2 + \cdots + e_n^2 \tag{4}$$

The least squares approach chooses $\hat{\beta_0}$ and $\hat{\beta_1}$ to minimize the RSS. The equation is as follows:

$$\hat{\beta_1} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{5}$$

$$\hat{\beta_0} = \bar{y} - \hat{\beta_1}\bar{x} \tag{6}$$

1.2 Assessing the Accuracy of the Coefficients Estimates

How accurate is the sample mean $\hat{\mu}$ as an estimate of μ? We answer this question by computing *standard error* of $\hat{\mu}$. We have the well known formula:

$$Var(\hat{\mu}) = SE(\hat{\mu})^2 = \frac{\sigma^2}{n} \tag{7}$$

Standard errors associated with the predicted coefficients is given as:

$$SE(\beta_0)^2 = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right] \tag{8}$$

$$SE(\beta_1)^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{9}$$

Here $\sigma^2 = Var(\epsilon)$. In general σ^2 is not known but can be estimated from the data. The estimate is known as residual standard error and is given by the formula:

$$RSE = \sqrt{RSS/(n-2)} \tag{10}$$

Standard error can be used to calculate confidence interval. A 95% confidence interval means the actual value lies in that interval with 95% probability.

For Linear Regression, a 95% confidence interval for $\hat{\beta}_1$ and $\hat{\beta}_0$ is given as:

$$\hat{\beta}_1 \pm 2 \cdot SE(\hat{\beta}_1) \tag{11}$$

$$\hat{\beta}_0 \pm 2 \cdot SE(\hat{\beta}_0) \tag{12}$$

Standard errors can also be used to perform *hypothesis tests* on the coefficients. The most common hypothesis test involves testing the *null hypothesis* of:

H_0 : There is no relationship between X and Y.

versus the *alternative hypothesis*

H_a: There is some relationship between X and Y.

If we take the equation:

$$Y = \beta_0 + \beta_1 X \tag{13}$$

Mathematically the *null hypothesis* corresponds to testing:

$$H_0 : \beta_1 = 0 \tag{14}$$

versus

$$H_a : \beta_1 \neq 0 \tag{15}$$

To test the null hypothesis, we need to verify that $\hat{\beta}_1$ is sufficiently far from zero. For this we use $SE(\hat{\beta}_1)$. If $SE(\hat{\beta}_1)$ is small

2 Multiple Linear Regression

This is the second section