Miscellaneous Notes on Regression

Based on SJS and KNN

Ⓒ Tingfeng Xia

Fall 2019, modified on November 8, 2019

Preface

Notes for STA302H1F fall offering, 2019 with Prof. Shivon Sue-Chee. These notes are based on the KNN and SJS text, in an aim for better understanding of the course material.

Contents

1 Weighted Least Square Regression

1.1 Motivation and Set-Up

Consider the straight line (simple) linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + e_i \quad \text{where} \quad e_i \sim N(0, \frac{\sigma^2}{w_i})$$

For the weight w_i, we should note the following

- $w_i \to \infty \implies Var(e_i) \to 0$. In this case, the estimates of the regression parameters β_0, β_1 should be such that the fitted line at x_i should be very close to y_i. (Small variance means more strict in terms of deviation from the regressin line, corresponding to a larger emphasis on the i-th data point.)

- If w_i is some small value, then the variance of the i-th data point would be quite large. In this case, we have a loose restriction of the deviation of the i-th data point from the regression line meaning that litte emphasis is taken for this data point.

- $w_i \to 0 \implies Var(e_i) \to \infty$. In this case, we have the variance tending to infinity. Meaning that there is absolutely no restriction/emphasis on the i-th data point and it could be simply removed from the set.

We define the cost function, WRSS as

$$\text{WRSS} = \sum_{i=1}^{n} w_i \left(y_i - \hat{y}_{W_i}\right)^2 = \sum_{i=1}^{n} w_i \left(y_i - b_0 - b_1 x_i\right)^2$$

and the estimators $\mathbf{b} = [b_0, b_1]^T$ are derived using MLE.

Intuition behind WRSS This cost function may seem wierd at first glance, but it intuitively makes sense. Notice that when w_i is large, the i-th lost term $w_i \left(y_i - \hat{y}_{W_i}\right)^2$ is payed more emphasis on. On the contrary, when $w_0 \to 0$, the term $\to 0$. (Indeed, when Variance of the term $\to \infty$ we just neglect it.)

1.2 Deriving ML Estimators

Derivatives

$$\frac{\partial \text{WRSS}}{\partial b_0} = -2 \sum_{i=1}^{n} w_i \left(y_i - b_0 - b_1 x_i\right) = 0 \tag{1}$$

$$\frac{\partial \text{WRSS}}{\partial b_1} = -2\sum_{i=1}^{n} w_i x_i \left(y_i - b_0 - b_1 x_i\right) = 0 \tag{2}$$

Normal Equations Obtained from rearranging the above equations,

$$\sum_{i=1}^{n} w_i y_i = b_0 \sum_{i=1}^{n} w_i + b_1 \sum_{i=1}^{n} w_i x_i \tag{3}$$

$$\sum_{i=1}^{n} w_i x_i y_i = b_0 \sum_{i=1}^{n} w_i x_i + b_1 \sum_{i=1}^{n} w_i x_i^2 \tag{4}$$

Rearranging Use (3) $\sum_{i=1}^{n} w_i x_i$ and (4) $\sum_{i=1}^{n} w_i$

$$\sum_{i=1}^{n} w_i x_i \sum_{i=1}^{n} w_i y_i = b_0 \sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_i + b_1 \left(\sum_{i=1}^{n} w_i x_i\right)^2 \tag{5}$$

$$\sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_i y_i = b_0 \sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_i + b_1 \sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_i^2 \tag{6}$$

WLS Slope Estimator [1]

$$\begin{aligned}
\hat{\beta}_{1W} &= \frac{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_i y_i - \sum_{i=1}^{n} w_i x_i \sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} \sum_{i=1}^{n} w_i x_i^2 - \left(\sum_{i=1}^{n} w_i x_i\right)^2} && (7)\\
&= \frac{\sum_{i=11}^{n} x_i \left(x_i - \bar{x}_W\right)\left(y_i - \bar{y}_W\right)}{\sum_{i=1}^{n} w_i \left(x_i - \bar{x}_W\right)^2} && (8)
\end{aligned}$$

WLS Intercept Estimator

$$\hat{\beta}_{0W} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} - \hat{\beta}_{1W} \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} = \bar{y}_w - \hat{\beta}_{1W} \bar{x}_W \tag{9}$$

[1] Note that $\bar{x}_W = \sum_{i=1}^{n} w_i x_i / \sum_{i=1}^{n} w_i$ and $\bar{y}_W = \sum_{i=1}^{n} w_i y_i / \sum_{i=1}^{n} w_i$