The University of Texas
at Austin

CS363D Statistical Learning and Data Mining

Homework 04

Edited by LaTeX

Department of Computer Science

STUDENT
Jimmy Lin
xl5224

INSTRUCTOR
Pradeep Ravikumar

TASSISTANT
Adarsh Prasad

RELEASE DATE
March. 25 2014

DUE DATE
April. 16 2014

TIME SPENT
7 hours

April 10, 2014

Contents

List of Figures

1 Variation of Linear Regression

In the scenario of one-dimensional input, we have input matrix $\mathbf{X}$ to be

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \quad \hat{\mathbf{X}} = \begin{pmatrix} 1 & \hat{x}_1 \\ 1 & \hat{x}_2 \\ \vdots & \vdots \\ 1 & \hat{x}_n \end{pmatrix} \quad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \tag{1}$$

According to the normal equation, we have

$$w^*_{(\mathbf{X})} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{2}$$

By (1), we have

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} 1 & \Sigma_{i=1}^n x_i \\ \Sigma_{i=1}^n x_i & \Sigma_{i=1}^n x_i^2 \end{pmatrix} \tag{3}$$

By inverting above equation, we have

$$(\mathbf{X}^T\mathbf{X})^{-1} = \frac{1}{\Sigma_{i=1}^n x_i^2 - (\Sigma_{i=1}^n x_i)^2} \begin{pmatrix} \Sigma_{i=1}^n x_i^2 & -\Sigma_{i=1}^n x_i \\ -\Sigma_{i=1}^n x_i & 1 \end{pmatrix} \tag{4}$$

And for representational convenience, we compute $\mathbf{X}^T\mathbf{Y}$ first as follows,

$$\mathbf{X}^T\mathbf{Y} = \begin{pmatrix} \Sigma_{i=1}^n y_i \\ \Sigma_{i=1}^n x_i y_i \end{pmatrix} \tag{5}$$

From above series of computatoin, we have

$$w^*_{0(\mathbf{X})} = w^*_{0(\mathbf{Y})} = \frac{1}{\Sigma_{i=1}^n x_i^2 - (\Sigma_{i=1}^n x_i)^2} \left(\Sigma_{i=1}^n x_i^2 \; \Sigma_{i=1}^n y_i - \Sigma_{i=1}^n x_i \; \Sigma_{i=1}^n x_i y_i\right) \tag{6}$$

$$w^*_{1(\mathbf{X})} = w^*_{1(\mathbf{X})} = \frac{1}{\Sigma_{i=1}^n x_i^2 - (\Sigma_{i=1}^n x_i)^2} \left(\Sigma_{i=1}^n x_i \; \Sigma_{i=1}^n y_i - \Sigma_{i=1}^n x_i y_i\right) \tag{7}$$

Similarly, for $\hat{\mathbf{X}}$, we have

$$w^*_{0(\hat{\mathbf{X}})} = \frac{1}{\Sigma_{i=1}^n \hat{x_i}^2 - (\Sigma_{i=1}^n \hat{x_i})^2} \left(\Sigma_{i=1}^n \hat{x_i}^2 \; \Sigma_{i=1}^n y_i - \Sigma_{i=1}^n \hat{x_i} \; \Sigma_{i=1}^n \hat{x_i} y_i\right) \tag{8}$$

$$w^*_{1(\hat{\mathbf{X}})} = \frac{1}{\Sigma_{i=1}^n \hat{x_i}^2 - (\Sigma_{i=1}^n \hat{x_i})^2} \left(\Sigma_{i=1}^n \hat{x_i} \; \Sigma_{i=1}^n y_i - \Sigma_{i=1}^n \hat{x_i} y_i\right) \tag{9}$$

And similar for $\hat{\mathbf{Y}}$, we have

$$w^*_{0(\hat{\mathbf{Y}})} = \frac{1}{\Sigma_{i=1}^n \hat{x_i}^2 - (\Sigma_{i=1}^n \hat{x_i})^2} \left(\Sigma_{i=1}^n x_i^2 \; \Sigma_{i=1}^n \hat{y_i} - \Sigma_{i=1}^n x_i \; \Sigma_{i=1}^n x_i \hat{y_i}\right) \tag{10}$$

$$w^*_{1(\hat{\mathbf{Y}})} = \frac{1}{\Sigma_{i=1}^n \hat{x_i}^2 - (\Sigma_{i=1}^n \hat{x_i})^2} \left(\Sigma_{i=1}^n x_i \; \Sigma_{i=1}^n \hat{y_i} - \Sigma_{i=1}^n x_i \hat{y_i}\right) \tag{11}$$

1.1 $\hat{x}_i = x_i - \bar{x}$

By taking $\hat{x}_i = x_i - \bar{x}$ and $\bar{x} = \frac{1}{n}\Sigma_{i=1}^n x_i$ into (8) and (9), we have

$$w^*_{0(\hat{\mathbf{X}})} = \Sigma_{i=1}^n y_i \neq w^*_{0(\mathbf{X})} \tag{12}$$

$$w^*_{1(\hat{\mathbf{X}})} = \frac{\Sigma_{i=1}^n x_i \ \Sigma_{i=1}^n y_i - \Sigma_{i=1}^n x_i y_i}{\Sigma_{i=1}^n x_i^2 - n\hat{x}_i{}^2} = w^*_{1(\mathbf{X})} \tag{13}$$

We will get **different w_0^* but the same w_1^***.

1.2 $\hat{x}_i = \alpha x_i$

$$\begin{aligned}
w^*_{0(\hat{\mathbf{X}})} &= \frac{\alpha^2\Sigma_{i=1}^n x_i^2 \ \Sigma_{i=1}^n y_i - \alpha^2\Sigma_{i=1}^n x_i \ \Sigma_{i=1}^n x_i y_i}{\alpha^2\Sigma_{i=1}^n x_i^2 - \alpha^2(\Sigma_{i=1}^n x_i)^2} && (14)\\
&= \frac{\Sigma_{i=1}^n x_i^2 \ \Sigma_{i=1}^n y_i - \Sigma_{i=1}^n x_i \ \Sigma_{i=1}^n x_i y_i}{\Sigma_{i=1}^n x_i^2 - (\Sigma_{i=1}^n x_i)^2} && (15)\\
&= w^*_{0(\mathbf{X})} && (16)\\
w^*_{1(\hat{\mathbf{X}})} &= \frac{\alpha\Sigma_{i=1}^n x_i^2 \ \Sigma_{i=1}^n y_i - \alpha\Sigma_{i=1}^n x_i \ \Sigma_{i=1}^n x_i y_i}{\alpha^2\Sigma_{i=1}^n x_i^2 - \alpha^2(\Sigma_{i=1}^n x_i)^2} && (17)\\
&= \frac{1}{\alpha}\frac{\Sigma_{i=1}^n x_i^2 \ \Sigma_{i=1}^n y_i - \Sigma_{i=1}^n x_i \ \Sigma_{i=1}^n x_i y_i}{\Sigma_{i=1}^n x_i^2 - (\Sigma_{i=1}^n x_i)^2} && (18)\\
&= \frac{1}{\alpha} w^*_{1(\mathbf{X})} && (19)
\end{aligned}$$

Hence, we can conclude that the w_0^* **remain unchanged and w_1^* becomes $\frac{1}{\alpha}$ of original**, in the case of $\hat{x}_i = \alpha x_i$.

1.3 $y_i = \alpha y_i$

$$\begin{aligned}
w^*_{0(\hat{\mathbf{Y}})} &= \frac{\alpha\Sigma_{i=1}^n x_i^2 \ \Sigma_{i=1}^n y_i - \alpha\Sigma_{i=1}^n x_i \ \Sigma_{i=1}^n x_i y_i}{\Sigma_{i=1}^n x_i^2 - (\Sigma_{i=1}^n x_i)^2} && (20)\\
&= \alpha\frac{\Sigma_{i=1}^n x_i^2 \ \Sigma_{i=1}^n y_i - \Sigma_{i=1}^n x_i \ \Sigma_{i=1}^n x_i y_i}{\Sigma_{i=1}^n x_i^2 - (\Sigma_{i=1}^n x_i)^2} && (21)\\
&= \alpha w^*_{0(\mathbf{Y})} && (22)\\
w^*_{1(\hat{\mathbf{Y}})} &= \frac{\alpha\Sigma_{i=1}^n x_i^2 \ \Sigma_{i=1}^n y_i - \alpha\Sigma_{i=1}^n x_i \ \Sigma_{i=1}^n x_i y_i}{\Sigma_{i=1}^n x_i^2 - (\Sigma_{i=1}^n x_i)^2} && (23)\\
&= \alpha\frac{\Sigma_{i=1}^n x_i^2 \ \Sigma_{i=1}^n y_i - \Sigma_{i=1}^n x_i \ \Sigma_{i=1}^n x_i y_i}{\Sigma_{i=1}^n x_i^2 - (\Sigma_{i=1}^n x_i)^2} && (24)\\
&= \alpha w^*_{1(\mathbf{Y})} && (25)\\
& && (26)
\end{aligned}$$

Hence, we can conclude that **the w_0^* and w_1^* increase to α multiples of original**.

2 Why not k-means?

Observation: entities of two different classes approximately form two circles of distinct radius.
Conclusion: k-means algorithm will not discover two clusters.
Explanation:

Logically, we provide the explicit explanation by two conditions. We first use notation s to denote the scenario of two clusters are precisely discovered by k-means algorithm.

Suppose that by execution of k-means algorithm, s is not reachable. Without any doubt, the two clusters cannot be precisely discovered in this case because we postulate that s is not reachable.

Suppose that s is reachable by k-means algorithm, s is not stable state (not the state to which algorithm will converged). Put it another way, given those two clusters ("o" and "+") are discovered, they will not after recomputation of centroid, and reassignment of data points (two stages in k-means algorithm). The centroids of two clusters will be in the center of the diagram, according to the recomputation of cluster centroid, which takes the means of all cluster members. Based on our observation above, both blue "o" and red "+" points will be segregated by new cluster membership after the reassignment of these data points.

3 Hierarchical Clustering

3.1 Single-Link Clustering

3.2 Complete-Link Clustering