ECE4710J Assignment 4

Name: Zhou Haoquan
ID:519370910059

March 2022

1 Properties of simple linear regression

(a)

We have

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{y}_i = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} (\bar{y} + r\sigma_y \frac{x - \bar{x}}{\sigma_x})$$
$$= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \bar{y} - r\sigma_y \frac{\sum_{i=1}^{n} x - \sum_{i=1}^{n} \bar{x}}{\sigma_x}$$

Note that $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \bar{x}$ and $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \bar{y}$, thus we get

$$\sum_{i=1}^{n} e_i = 0 - r\sigma_y \frac{0}{\sigma_x} = 0$$

(b)

We have proved that

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{y}_i = 0$$

we devide the equation by n and get

$$\frac{1}{n} \sum_{i=1} y_i - \frac{1}{n} \sum_{i=1}^{n} \hat{y}_i = 0 \Rightarrow \bar{y} - \bar{\hat{y}} = 0 \Rightarrow \bar{y} = \bar{\hat{y}}$$

(c)

We plug $\bar{x}$ into the simple linear regression equation

$$\hat{y}|_{x=\bar{x}} = \bar{y} + r\sigma_y \frac{\bar{x} - \bar{x}}{\sigma_x} = \bar{y}$$

Thus $(\bar{x}, \bar{y})$ is on the regression line.

2 Geometric Perspective of Least Square

(a)

We consider the orthogonal properties ($e \perp span(\mathbb{I}, \vec{x}) \Rightarrow e \perp \mathbb{I}$ and $e \perp \vec{x}$) and write:

$$\hat{\theta_1}\vec{x} \perp \mathbb{Y} - \hat{\theta_1}\vec{x}$$

and

$$\hat{\theta_0} \perp \mathbb{Y} - \theta_0$$

Expand them and write

$$\sum_{i=1}^{n} \hat{\theta_{0,i}}(y_i - \hat{\theta_{0,i}}) = 0 \tag{1}$$

$$\sum_{i=1}^{n} \hat{\theta_{1,i}}x_i(y_i - \hat{\theta_{1,i}}x_i) = 0 \tag{2}$$

we substract (2) from (1) and get

$$\sum_{i=1}^{n}[(\hat{\theta_{0,i}} - \hat{\theta_{1,i}}x_i)y_i - (\hat{\theta_{0,i}}^2 - (\hat{\theta_{1,i}}x_i)^2)] = 0$$

$$\Rightarrow (\hat{\theta_{0,i}} - \hat{\theta_{1,i}}x_i)[\sum_{i=1}^{n}(y_i - (\hat{\theta_{0,i}} + \hat{\theta_{1,i}}x_i))] = 0$$

Since $\mathbb{I}$ and $\vec{x}$ is linearly independent (otherwise this is not a intercept), we can deduce that

$$\sum_{i=1}^{n}(y_i - (\hat{\theta_{0,i}} + \hat{\theta_{1,i}}x_i)) = 0 \Rightarrow \sum_{i=1}^{n} e_i = 0$$

(b)

We start with the fact $e \perp \mathbb{X}\theta$, and write this relation into

$$e(\hat{\theta_0} + \hat{\theta_1}\vec{x})^T = 0$$

split it

$$e\hat{\theta_0}^T + e(\hat{\theta_1}\vec{x})^T = 0$$

Note that both terms above must be 0, otherwise

$$e\hat{\theta_0}^T = -e(\hat{\theta}_1\vec{x})^T \neq 0$$

which indicates that

$$\hat{\theta}_0 = -\hat{\theta}_1\vec{x}$$

This contradicts to the fact that $\mathbb{I}$ is a intercept term. Thus we have

$$e(\hat{\theta}_1\vec{x}) = 0 \Rightarrow e \perp \hat{\theta}_1\vec{x} \Rightarrow e \perp \vec{x}$$

(c)

We want to minimize the $e = \mathbb{Y} - \hat{\mathbb{Y}}$. We may consider e is not orthogonal to $\hat{Y}$, then we can find a vector $\vec{v} \in \text{span}(\mathbb{I}, \vec{x})$ such that

$$\mathbb{Y} = \hat{\mathbb{Y}} + \vec{v} - \vec{v} + e$$

We now adjust $\vec{v}$ to make $e - \vec{v}$ the shortest. Note that $\hat{Y} + \vec{v} \in \text{span}(\mathbb{I}, \vec{x})$. In this way, we construct a better $\hat{\mathbb{Y}}' = \hat{Y} + \vec{v}$ with new residual vector $e^{'} = e - \vec{v}$, shown in the following figure

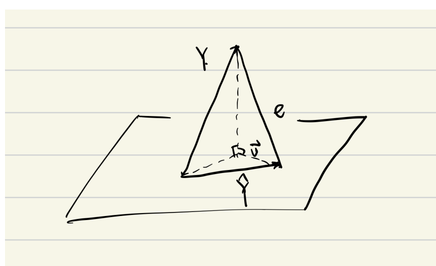

Figure 1: A new constructed residual vector that has smaller length

3 Properties of a Linear Model With No Constant Term

Let

$$R^{'}(\gamma) = \frac{1}{n}\sum_{i=1}^{n}(-2x_i)(y_i - \gamma x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} -x_i y_i + \sum_{i=1}^{n} x_i^2 \gamma = 0 \Rightarrow \hat{\gamma} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_1^2}$$

We now prove this is the global minimum:

$$R''(\gamma) = \sum_{i=1}^{n} 2x_i^2 > 0$$

Thus this is a global minimum.

4

(a)

This is no false. We may perform similar operation:

$$(\mathbb{Y} - \gamma x)\ \bar{}x \Rightarrow \sum_{i=1}^{n} [(y_i - \gamma x_i) xi] = 0$$

This can be further simplified to

$$\sum_{i=1}^{n} y_i x_i - \gamma \sum_{i=1}^{n} x_i^2 = 0$$

In general case,

$$\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} \gamma x_i^2 = 0 \not\Leftrightarrow \sum_{i=1}^{n} y_i - \sum_{i=0}^{n} \gamma x_i = 0$$

(b)

This is true. To make e minimum, we need to make $(\mathbb{Y} - \gamma x) \perp \gamma x$. And this can be done by projecting $\mathbb{Y}$ to the line expanded by x, namely $\mathrm{span}(x)$. And

$$(\mathbb{Y} - \gamma x) \perp \gamma x \Rightarrow (\mathbb{Y} - \gamma x) \perp x \Rightarrow e \perp x$$

(c)

This is true. As we discussed before, we need to project $\mathbb{Y}$ to $\mathrm{span}(x)$ in order to get the minimum e. And the projecting result is $\hat{\mathbb{Y}} = \gamma x$, which is orthogonal to e.

(d)

This is false. We can look to the result of (a). We conclude that in general

$$\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} \gamma x_i^2 = 0 \not\Leftrightarrow \sum_{i=1}^{n} y_i - \sum_{i=0}^{n} \gamma x_i = 0$$

We find that the later statement is just

$$\bar{y} = \gamma \bar{x} \tag{3}$$

which means $(\bar{x}, \bar{y})$ is on the regression line. Howwever, we see that we cannot draw (3) from current regression model. Thus, $(\bar{x}, \bar{y})$ is not on the regression line in general case.

5 MSE "Minimizer"

(a)

quadratic, γ

(b)

The first derivative is

$$g_i^{'}(\gamma) = \frac{-2x_i}{n}(y_i - \gamma x_i)$$

The second derivative is

$$g_i^{''}(\gamma) = \frac{2x_i^2}{n} \geq 0$$

Thus we can conclude that g_i is convex.

(c)

Since, the function is convex, the whole function is facing upwards. And we need to notice that the function must be monotonically decreasing when $x < x_0$ (x_0 is the critical point), and monotonically increasing when $x > x_0$. otherwise, there will be a region where the function will actually become concave. Like the following figure shows Thus, by solving $\frac{dg(x)}{dx} = 0$, we can get the the critical point which minimizes $g(x)$.

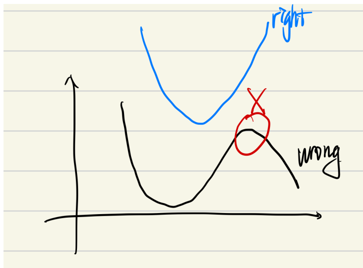

Figure 2: A convex function

(d)

(i)

Define
$$f(x) = g(x) + h(x)$$
Then
$$f(cx_1 + (1-c)x_2) = g(cx_1 + (1-c)x_2) + h(cx_1 + (1-c)x_2)$$
Since both $g(x)$ and $h(x)$ are convex functions, we can apply the inequality and get

$$\begin{aligned} f(cx_1 + (1-c)x_2) &= g(cx_1 + (1-c)x_2) + h(cx_1 + (1-c)x_2) \\ &\leq cg_(x_1) + (1-c)g(x_2) + ch(x_1) + (1-c)h(x_2) \\ &= cf(x_1) + (1-c)f(x_2) \end{aligned}$$

Thus $g(x) + h(x)$ is also a convex function.

(ii)

As we discussed above, two convex function sums up to one convex function. And this new convex function can again be summed up with another convex function to get a even newer convex function. By doing this summation for $n-1$ times, we can prove that the sum of n convex function is still a convex function.

(e)

Since all the g_is are convex function (proved in (b)), we can say that all of them will sum up to a new convex function (proved in (d)). And we proved in (c) that the critical point for a convex function is the global minimum. Thus we can conclude that the critical point of MSE will mimimize the MSE.