

# CS 760: Machine Learning - Fall 2020

## Homework 2: Linear Regression

Due : 10/13/2020

Zijie Zhang

October 10, 2020

### Problem 1

*Proof.* We notice that:

$$\begin{aligned}\|\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\|_2^2 &= (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) \\ &= \mathbf{y}^T \mathbf{y} - 2\boldsymbol{\theta}^T \mathbf{X}^T \mathbf{y} + \boldsymbol{\theta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\theta}\end{aligned}$$

Compute the differential:

$$\begin{aligned}d \operatorname{tr} (\mathbf{y}^T \mathbf{y} - 2\boldsymbol{\theta}^T \mathbf{X}^T \mathbf{y} + \boldsymbol{\theta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\theta}) &= -2 \operatorname{tr} ((d\boldsymbol{\theta})^T \mathbf{X}^T \mathbf{y}) + \operatorname{tr} ((d\boldsymbol{\theta})^T \mathbf{X}^T \mathbf{X} \boldsymbol{\theta}) + \operatorname{tr} (\boldsymbol{\theta}^T \mathbf{X}^T \mathbf{X} (d\boldsymbol{\theta})) \\ &= 2 \operatorname{tr} ((d\boldsymbol{\theta})^T (\mathbf{X}^T \mathbf{X} \boldsymbol{\theta} - \mathbf{X}^T \mathbf{y}))\end{aligned}$$

It follows that

$$\frac{d}{d\boldsymbol{\theta}} = 2\mathbf{X}^T \mathbf{X} \boldsymbol{\theta} - 2\mathbf{X}^T \mathbf{y}$$

Let this derivative to zero, we have

$$\arg \min_{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

□

### Problem 2

*Proof.* We know that

$$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\theta}^*, \boldsymbol{\Sigma}^*)$$

Then

$$\mathbb{P}(\mathbf{y}, \mathbf{X} | \boldsymbol{\theta}, \boldsymbol{\Sigma}^*) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{\Sigma}^*|}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T \boldsymbol{\Sigma}^{*-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) \right\}$$

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg \max_{\boldsymbol{\theta}} \mathbb{P}(\mathbf{y}, \mathbf{X} | \boldsymbol{\theta}, \boldsymbol{\Sigma}^*) \\ &= \arg \max_{\boldsymbol{\theta}} \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{\Sigma}^*|}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T \boldsymbol{\Sigma}^{*-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) \right\} \\ &= \arg \max_{\boldsymbol{\theta}} -(\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T \boldsymbol{\Sigma}^{*-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) \\ &= \arg \min_{\boldsymbol{\theta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T \boldsymbol{\Sigma}^{*-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})\end{aligned}$$

Compute the differential:

$$\begin{aligned}\frac{d}{d\boldsymbol{\theta}} &= 2\mathbf{X}^T \boldsymbol{\Sigma}^{*-1} \mathbf{X} \boldsymbol{\theta} - 2\mathbf{X}^T \boldsymbol{\Sigma}^{*-1} \mathbf{y} \\ \hat{\boldsymbol{\theta}} &= (\mathbf{X}^T \boldsymbol{\Sigma}^{*-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{*-1} \mathbf{y}\end{aligned}$$

□

### Problem 3

*Proof.*

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}\left(\boldsymbol{\theta}^*, |\boldsymbol{\Sigma}^*|^{\frac{1}{N}} \left(\mathbf{X}^T \boldsymbol{\Sigma}^{*-1} \mathbf{X}\right)^{-1}\right)$$

□

### Problem 4

*Proof.*

$$\hat{\mathbf{y}} = \mathbf{x}^T \hat{\boldsymbol{\theta}} = \mathbf{x}^T \left(\mathbf{X}^T \boldsymbol{\Sigma}^{*-1} \mathbf{X}\right)^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{*-1} \mathbf{y}$$

□

### Problem 5

*Proof.*

$$\hat{\mathbf{y}} \sim \mathcal{N}\left(\mathbf{x}^T \boldsymbol{\theta}^*, |\boldsymbol{\Sigma}^*|^{\frac{1}{N}} \mathbf{x}^T \left(\mathbf{X}^T \boldsymbol{\Sigma}^{*-1} \mathbf{X}\right)^{-1} \mathbf{x}\right)$$

□

### Problem 6

*Proof.*

$$\begin{aligned} \hat{\boldsymbol{\Sigma}} &= \arg \max_{\boldsymbol{\Sigma}} \mathbb{P}(\mathbf{y}, \mathbf{X} | \boldsymbol{\theta}^*, \boldsymbol{\Sigma}) \\ &= \arg \max_{\boldsymbol{\Sigma}} \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{\Sigma}^*|}} \exp\left\{-\frac{1}{2N} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T \boldsymbol{\Sigma}^{*-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})\right\} \\ &= -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log(|\boldsymbol{\Sigma}^*|) - \frac{1}{2N} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T \boldsymbol{\Sigma}^{*-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) \end{aligned}$$

Taking the derivative we have:

$$\frac{d}{d\sigma} \log(\mathbb{P}(\mathbf{y}, \mathbf{X} | \boldsymbol{\theta}^*, \boldsymbol{\Sigma})) = \left( \boldsymbol{\Sigma}^{*-1} - \frac{1}{N} \boldsymbol{\Sigma}^{*-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}^*)^T (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}^*) \boldsymbol{\Sigma}^{*-1} \right)^T$$

Setting to zero,

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}^*)^T (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}^*)$$

□

### Problem 7

*Proof.*

(a)

$$\begin{aligned} \boldsymbol{\Sigma} &= \begin{bmatrix} 179.6875 & 68.75 \\ 68.75 & 487.5 \end{bmatrix} \\ \boldsymbol{\theta}^* &= \begin{bmatrix} -0.11278259 \\ 0.99759605 \end{bmatrix} \end{aligned}$$

(b)

$$\hat{\mathbf{y}} = \mathbf{x}^T \hat{\boldsymbol{\theta}} = 149.85437495$$

(c)

(d) No. Since the parameter related to height is close to 0.

(e) Yes. Since the parameter related to weight is close to 1.

□