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:

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

Compute the differential:

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

It follows that

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

Let this derivative to zero, we have

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

□

Problem 2

Proof. We know that

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

Then

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

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

Compute the differential:

$$\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}} = 2\mathbf{X}^T\boldsymbol{\Sigma}^{*-1}\mathbf{X}\boldsymbol{\theta} - 2\mathbf{X}^T\boldsymbol{\Sigma}^{*-1}y$$

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

□

Problem 3

Proof.

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

□

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.

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

□

Problem 6

Proof.

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

Taking the derivative we have:

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

Setting to zero,

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{N}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\theta}\right)^T\left(\mathbf{y}-\mathbf{X}\boldsymbol{\theta}\right)$$

□

Problem 7

Proof.

(a)

(b)

(c)

(d)

(e)

□