CS 760: Machine Learning Spring 2024

Homework 2: Linear Regression

Authors: Jed Pulley

<u>DO NOT POLLUTE!</u> AVOID PRINTING, OR PRINT 2-SIDED MULTIPAGE.

Problem 2.1

To minimize the MSE of θ, we computer our estimator $\hat{\theta}$:

$$\hat{\theta} = \arg\min_{\theta} \|\mathsf{y} - \mathbf{X}\theta\|_{\mathbf{2}}^{\mathbf{2}}$$

Expanding that out, we find:

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

From there, we take the derivative w.r.t. θ:

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

After setting this derivative to zero and solving for θ, we find that our solution is:

$$\hat{\theta} = (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-\mathbf{1}}\mathbf{X}^{\mathsf{T}}\mathsf{y}$$

Problem 2.2

Since know that ϵ is distributed normally, we can use the general multivariate normal distribution formula to derive our expression for the MLE of θ^*:

$$\mathbb{P}(\mathsf{y}, \mathbf{X}|\theta, \mathbf{\Sigma}^*) = \frac{\mathbf{1}}{\mathbf{2}\pi^{\frac{\mathbf{n}}{\mathbf{2}}}|\mathbf{\Sigma}^*|}\mathbf{e}^{-\frac{\mathbf{1}}{\mathbf{2}}(\mathsf{y}-\mathbf{X}\theta)^{\mathbf{T}}\mathbf{\Sigma}^{*-\mathbf{1}}(\mathsf{y}-\mathbf{X}\theta)}$$

To find the MLE, we first take the natural log, then we take the derivative and set it to zero. After that, we solve for $\hat{\theta}$:

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

This is very similar to the solution of Problem 2.1 above, but it's weighted by the inverse covariance matrix Σ^*

Problem 2.3

Problem 2.4

Problem 2.5

Problem 2.6

Problem 2.7

(a)