

Homework 2: Linear Regression

AUTHORS: Jed Pulley

DO NOT POLLUTE! AVOID PRINTING, OR PRINT 2-SIDED MULTIPAGE.**Problem 2.1**

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

$$\hat{\theta} = \arg \min_{\theta} \|y - \mathbf{X}\theta\|_2^2$$

Expanding that out, we find:

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

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

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

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

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

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^* :

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

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 $\hat{\theta}$:

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

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

Thanks to the central limit theorem, we can assume that θ^* is distributed normally under large sample sizes. The mean of this distribution is found to be θ^* , but our covariance matrix is a little trickier to find. We need to first compute the Fisher Information matrix, given by:

$$I(\theta^*) = -\mathbb{E}\left[\frac{\partial^2}{\partial \theta^2} \ln(\mathbb{P}(\mathbf{y}, \mathbf{X} | \theta, \Sigma^*))\right]$$

Therefore, the distribution of θ^* can be shown as:

$$\theta^* \sim \mathcal{N}(\theta^*, I(\theta^*)^{-1})$$

Problem 2.4

Problem 2.5

Problem 2.6

Problem 2.7

(a)