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 computer our estimator $\hat{\theta}$:

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

Expanding that out, we find:

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

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

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

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

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

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^*:

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

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 $\hat{\theta}$:

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

This is very similar to the solution of Problem 2.1 above, with the exception that it's now being weighted by the inverse covariance matrix $\mathbf{\Sigma}^{*-1}$

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:

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

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

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

Problem 2.4

Given a new feature vector x, our new response is: $\hat{y} = x^T\hat{\theta}$. Letting $H = (\mathbf{X}^{\mathsf{T}}\mathbf{\Sigma}^{*-1}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{\Sigma}^{*-1}$, given the above equation for $\hat{\theta}$, we can show that the MLE of our new response is $\hat{y} = x^T\theta^* + x^T H\epsilon$

Problem 2.5

$\hat{y} \sim \mathcal{N}(x^T\theta^*, x^T H\Sigma^* H^T x)$

Problem 2.6

Problem 2.7

(a)