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^*:

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

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 $\hat{\theta}$:

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

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 Σ^{*-1}

Problem 2.3

The distribution of the MLE of θ^* is:

$$\hat{\theta} \sim \mathcal{N}(\theta^*, (X^T \Sigma^{*-1} X)^{-1})$$

Problem 2.4

With x being a new sample and $\hat{\theta}$ being the MLE of θ^*, we can say that based on what we derived above, the MLE of $\hat{y}$ is exactly $x^T \hat{\theta}$

Problem 2.5

Borrowing from 2.3 and knowing that $\hat{y} = x^T \hat{\theta}$, we say that:

$$\hat{y} \sim \mathcal{N}(x^T \theta^*, x^T (X^T \Sigma^{*-1} X)^{-1} x)$$

Problem 2.6

We'll take a similiar approach as in Problem 2.2, so first we take the natural log, take the derivative w.r.t. Σ^*, set it to zero, and then solve for Σ^*. Doing so, we get:

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

Problem 2.7

(a) Here we run into an issue, because both Σ^* and θ^* are defined in terms of each other. To get around this issue, we use Expectation Maximization to iteratively approach the solution for either terms. Doing this yields an MLE for $\Sigma^* = 1025.0$ and $\theta^* = 155$

(b) Following the same logic as Problem 2.4, the MLE of $\hat{y}$ is $x^T \hat{\theta}$

(c) To derive a confidence interval of 95%, we first need to find tau:

$$\tau = \Phi_{\mathcal{N}}^{-1}(\frac{\alpha}{2}|0, \sigma^2 \mathsf{x}^\mathsf{T}(\mathsf{X}^\mathsf{T}\mathsf{X})^{-1}\mathsf{x})$$

such that

$$\mathbb{P}(|\hat{y} - \mathsf{y}^*| \leq \tau) = 1 - \alpha$$

Here we set $\alpha = 0.05$ to give us a 95% confidence interval $(\hat{y} - \tau, \hat{y} + \tau)$

(d) No

(e) Yes