

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

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

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, 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 similar 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:

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

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:

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

such that

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

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

- (d) No. It did not seem to have a big enough effect on the coefficients
- (e) Yes. Inversely, this does seem to have a large effect on the coefficients