Homework 2: Linear Regression

Instructor: Daniel L. Pimentel-Alarcón Due 10/13/2020

SOLUTION

GO GREEN. AVOID PRINTING, OR PRINT 2-SIDED MULTIPAGE.

In class we studied the linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\theta}^\star + \boldsymbol{\epsilon}$, under the *homoskedastic* assumption $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. In this homework you will derive the same results for the slightly more general *heteroskedastic* model where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}^\star)$. Each subproblem is worth 10 points.

Problem 2.1. Derive an expression for the coefficient vector $\boldsymbol{\theta}$ that minimizes the mean squared error, i.e.,

$$\arg\min_{\boldsymbol{\theta}} \ \|\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\|_2^2.$$

Solution. This is the same as in the homoskedastic case, $\hat{\boldsymbol{\theta}} = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\mathbf{y}$.

Problem 2.2. Derive an expression for the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}^\star$, i.e.,

$$\arg\max_{\boldsymbol{\theta}} \ \mathbb{P}(\mathbf{y}, \mathbf{X}|\boldsymbol{\theta}, \boldsymbol{\Sigma}^\star)$$

Solution. For this, write:

$$\begin{aligned}
\hat{\boldsymbol{\theta}} &= \arg\max_{\boldsymbol{\theta}} \ \mathbb{P}(\mathbf{y}, \mathbf{X}|\boldsymbol{\theta}, \boldsymbol{\Sigma}^\star) \\
&= \arg\max_{\boldsymbol{\theta}} \ \frac{1}{\sqrt{(2\pi)^{\mathrm{N}}|\boldsymbol{\Sigma}^\star|}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\theta})^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\theta})} \\
&= \arg\max_{\boldsymbol{\theta}} \ -(\mathbf{y}-\mathbf{X}\boldsymbol{\theta})^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\theta}) \\
&= \arg\min_{\boldsymbol{\theta}} \ \mathbf{y}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\mathbf{y} - \mathbf{y}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\mathbf{X}\boldsymbol{\theta} - \boldsymbol{\theta}^\mathsf{T}\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\mathbf{y} + \boldsymbol{\theta}^\mathsf{T}\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\mathbf{X}\boldsymbol{\theta} \\
&= \arg\min_{\boldsymbol{\theta}} \ -2\boldsymbol{\theta}^\mathsf{T}\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\mathbf{y} + \boldsymbol{\theta}^\mathsf{T}\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\mathbf{X}\boldsymbol{\theta}.
\end{aligned}$$

Taking derivative with respect to $\boldsymbol{\theta}$ and setting to zero, we have:

$$-2\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\mathbf{y} + 2\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\mathbf{X}\boldsymbol{\theta} \ = \ \mathbf{0},$$

and solving we obtain:

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

Problem 2.3. What is the distribution of the MLE of $\boldsymbol{\theta}^\star$?

Solution. Since $\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\theta}^\star, \boldsymbol{\Sigma}^\star)$,

$$\begin{aligned}\hat{\boldsymbol{\theta}} &\sim \mathcal{N}\Big((\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\mathbf{X}\boldsymbol{\theta}^\star, (\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\boldsymbol{\Sigma}^\star((\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1})^\mathsf{T}\Big)\\ &= \mathcal{N}\Big(\boldsymbol{\theta}^\star, (\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\mathbf{X}(\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\mathbf{X})^{-1}\Big)\\ &= \mathcal{N}\Big(\boldsymbol{\theta}^\star, (\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\mathbf{X})^{-1}\Big).\end{aligned}$$

Problem 2.4. Given a new sample with feature vector $\mathbf{x}$, what is the MLE of the response, $\hat{y}$?

Solution. By the invariance property of the MLE, $\hat{y} = \mathbf{x}^\mathsf{T}\hat{\boldsymbol{\theta}}$.

Problem 2.5. Given a new sample with feature vector $\mathbf{x}$, what is the distribution of the MLE $\hat{y}$?

Solution. Since $\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}^\star, (\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\mathbf{X})^{-1})$, we have that $\hat{y} \sim \mathcal{N}(\mathbf{x}\boldsymbol{\theta}^\star, \mathbf{x}(\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}^{\star-1}\mathbf{X})^{-1}\mathbf{x}^\mathsf{T})$.

Problem 2.6. Derive an expression for the MLE of $\boldsymbol{\Sigma}^\star$, i.e.,

$$\arg\max_{\boldsymbol{\Sigma}} \ \mathbb{P}(\mathbf{y}, \mathbf{X}|\boldsymbol{\theta}^\star, \boldsymbol{\Sigma})$$

Solution. For this, write:

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\max_{\boldsymbol{\Sigma}} \ \mathbb{P}(\mathbf{y}, \mathbf{X}|\boldsymbol{\theta}, \boldsymbol{\Sigma})\\ &= \arg\max_{\boldsymbol{\Sigma}} \ \frac{1}{\sqrt{(2\pi)^{\mathrm{N}}|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\theta})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\theta})}\\ &= \arg\max_{\boldsymbol{\Sigma}} \ -\frac{\mathrm{N}}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - (\mathbf{y}-\mathbf{X}\boldsymbol{\theta})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\theta})\\ &= \arg\min_{\boldsymbol{\Sigma}} \ \frac{1}{2}\log|\boldsymbol{\Sigma}| + (\mathbf{y}-\mathbf{X}\boldsymbol{\theta})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\theta})\end{aligned}$$

Taking derivative with respect to $\boldsymbol{\Sigma}$ and setting to zero, we have:

$$\begin{aligned}\frac{1}{|\boldsymbol{\Sigma}|}|\boldsymbol{\Sigma}|\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\theta})(\mathbf{y}-\mathbf{X}\boldsymbol{\theta})^\mathsf{T}\Sigma^{-1} &= \mathbf{0}\\ \mathbf{I} - \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\theta})(\mathbf{y}-\mathbf{X}\boldsymbol{\theta})^\mathsf{T} &= \mathbf{0},\end{aligned}$$

and solving we obtain:

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

Problem 2.7. Consider the following vector $\mathbf{y}$, containing information about glucose level of four individuals, and the following data matrix $\mathbf{X}$ containing information about height and weight of the corresponding individuals:

$$\mathbf{y} = \begin{bmatrix}110\\140\\180\\190\end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix}180 & 150\\150 & 175\\170 & 165\\185 & 210\end{bmatrix}.$$

Given these data

(a) What are your maximum likelihood estimates of $\mathbf{\Sigma}^\star$ and $\boldsymbol{\theta}^\star$?

(b) Given a new sample with feature vector $\mathbf{x} = [175 \ \ 170]^\mathsf{T}$, what is the maximum likelihood estimate of its response $\hat{\mathrm{y}}$?

(c) Derive a 95% confidence interval for $\hat{\mathrm{y}}$.

(d) Would you conclude that height is a significant feature for this model? Why?

(e) Would you conclude that weight is a significant feature for this model? Why?