Practice problems

June 1, 2024

1. Suppose we want to transform the vector $\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$ to $v = \begin{bmatrix} v_1 \\ 0 \\ 0 \end{bmatrix}$. What is the householder matrix? (Fractions can appear! 분수가 나올 수 있음.)

2. Let $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$, $X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1N} & x_{2N} & \cdots & x_{pN} \end{bmatrix}$, $\mathcal{E} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$. Suppose we have

$$Y = X\beta + \mathcal{E}$$

where $\varepsilon_1, \cdots, \varepsilon_N$ are independent with $E[\varepsilon_i] = 0$ and $Var[\varepsilon_i] = \sigma^2$ for every i. Assume $N \geq p + 1$ and X has full rank. Prove that

$$\mathbb{E}\left[(Y - X\hat{\beta})^T(Y - X\hat{\beta})\right] = (N - p - 1)\sigma^2,$$

where $\hat{\beta}$ is the least-squares estimate.

3. Let $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$, $X = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{1N} & x_{2N} \end{bmatrix}$, $\mathcal{E} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$. Suppose we have

$$Y = X\beta + \mathcal{E}$$

where $\varepsilon_1, \cdots, \varepsilon_N \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$. Suppose also that the columns of X are **orthogonal** (but they are not unit vectors.) Finally, suppose that $N = 15$ and

$$\sum_{i=1}^{N} y_i = -30, \sum_{i=1}^{N} y_i^2 = 108, \sum_{i=1}^{N} x_{1i}^2 = 16, \sum_{i=1}^{N} x_{2i}^2 = 5, \sum_{i=1}^{N} x_{1i}y_i = 16, \sum_{i=1}^{N} x_{2i}y_i = 10.$$

(a) What is the R^2 of the least-squares estimate $\hat{\beta}$? (*Hint*: $R^2 = SSR/SST$) Explain your response.

(b) Suppose we want to conduct the following test.

$$H_0: \ \beta_1 \leq \frac{1}{2} \ \ vs \ \ H_1: \ \beta_1 > \frac{1}{2}$$

What is the Z-statistic for this test? Explain your response.

(c) What is the p-value for the test? Use the following table for the CDF of the standard normal distribution. Explain your response.

v	$P(Z < v)$
-3	0.0013
-2	0.0228
-1	0.1587
0	0.5

4. Consider the left dataset below with 2 variables (v_1 and v_2) and 3 samples. Suppose we do principal components analysis on this dataset based on the sample variance. (In this exercise, you don't need to standardize the dataset before doing PCA.)

	v_1	v_2
1	-1	1
2	0	0
3	1	1

$\Rightarrow$

	$PC1$
1	?
2	?
3	?

(a) What is the first principal component? (the answer is not unique) Show your derivation.

(b) How much of the total variance does the $PC1$ explains? Explain in proportion. Show your derivation.

5. Suppose we have $X = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{p1} \\ x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1N} & x_{2N} & \cdots & x_{pN} \end{bmatrix}$. Suppose we do spectral decomposition

$$X^T(I - \frac{1}{N}\mathbf{1}\mathbf{1}^T)X = V\Lambda V^T \quad (V, \Lambda \in \mathbb{R}^{p\times p},\ V^TV = VV^T = I \in \mathbb{R}^{p\times p},\ \Lambda : \text{diagonal})$$

where I in the left hand side is $N \times N$ identity matrix and $\mathbf{1} = [1, 1, \cdots, 1]^T \in \mathbb{R}^N$. Suppose

$$\tilde{X} = \begin{bmatrix} | & | & \cdots & | \\ \tilde{X}^1 & \tilde{X}^2 & \cdots & \tilde{X}^p \\ | & | & \cdots & | \end{bmatrix} = (I - \frac{1}{N}\mathbf{1}\mathbf{1}^T)XV$$

(a) Prove that columns of $\tilde{X}$ are orthogonal. (You just need to show that $\tilde{X}^i$ and $\tilde{X}^j$ are orthogonal when $i \neq j$.)

(b) Prove that $\mathbf{1}$ and columns of $\tilde{X}$ are orthogonal each other.

(c) Suppose we assume $y_i = \beta_0 + \beta_1\tilde{X}^1_i + \beta_2\tilde{X}^2_i + \varepsilon_i$ where ε_i are independent, have mean 0, and have variance $Var(\varepsilon_i) = \sigma^2$ $(i = 1, 2, \cdots, N)$. Prove that the least-squares estimates of β_1 and β_2 are

$$\hat{\beta}_1 = \frac{1}{\lambda_1}\sum_{i=1}^{N}\tilde{X}^1_i y_i, \quad \hat{\beta}_2 = \frac{1}{\lambda_2}\sum_{i=1}^{N}\tilde{X}^2_i y_i,$$

where $\lambda_1, \lambda_2, \cdots, \lambda_p$ are diagonal elements of Λ.

6. Maximum likelihood estimates. (In the exercises below, log-likelihoods have gradient zero at maximizers only.)

(a) Suppose we sampled $X_1, X_2, \cdots, X_5$ independently from a distribution with probability mass function $f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$ where $x = 0, 1, 2, ...$ and $\lambda > 0$. Suppose $(X_1, X_2, X_3, X_4, X_5) = (x_1, x_2, x_3, x_4, x_5)$. What is the MLE of λ? Explain your response.

(b) Suppose we sampled $X_1, X_2, \cdots, X_5$ independently from a distribution with the following probability mass function where $0 < \theta < 1$. Suppose $(X_1, X_2, X_3, X_4, X_5) = (0, 1, 1, 2, 2)$. What is the MLE of θ? Explain your response.

X	0	1	2	3
$p(X)$	$\frac{1}{4}\theta$	$\frac{3}{4}\theta$	$\frac{1}{4}(1-\theta)$	$\frac{3}{4}(1-\theta)$

7. Let $Y = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$, $X = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & 0 \\ 1 & 1 & 1 \end{bmatrix}$, $\mathcal{E} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$. Suppose we have

$$Y = X\beta + \mathcal{E}$$

where $\varepsilon_1, \varepsilon_2, \varepsilon_3$ are independent with $E[\varepsilon_i] = 0$ and $Var[\varepsilon_i] = \sigma^2$ for every i.

(a) Suppose we want to find the least-squares estimate of β by minimizing the loss function $\ell(\beta) = (Y - X\beta)^T(Y - X\beta)$ using the gradient descent algorithm with learning rate $\gamma = 0.1$. When $\beta^{(t)} = [1, \frac{1}{6}, -\frac{1}{2}]^T$, what is $\beta^{(t+1)}$? Explain your response.

(b) Suppose we want to find the least-squares estimate by minimizing the loss function $\ell(\beta) = (Y - X\beta)^T(Y - X\beta)$ using the Newton-Raphson algorithm. When $\beta^{(t)} = [1, \frac{1}{6}, -\frac{1}{2}]^T$, what is $\beta^{(t+1)}$? Explain your response.

8. (In this exercise, assume $\exp(1) = 3$.) Suppose we observe 32 pairs (x_1, y_1), (x_2, y_2), $\cdots$, (x_{32}, y_{32}) where

$$\begin{aligned} \text{for } 1 \le i \le 16, &\quad (x_i, y_i) = (1, 1) \\ \text{for } 17 \le i \le 20, &\quad (x_i, y_i) = (0, 1) \\ \text{for } 21 \le i \le 32, &\quad (x_i, y_i) = (0, 0) \end{aligned}$$

Suppose $y_1, y_2, \cdots, y_{32}$ are independent random variables and

$$y_i \sim Ber\left(\frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}\right)$$

where $\beta = [\beta_0, \beta_1]^T \in \mathbb{R}^2$.

(a) Suppose we want to find the MLE of β using the gradient descent algorithm with learning rate $\gamma = 0.1$. When $\beta^{(t)} = [0, 1]^T$, what is $\beta^{(t+1)}$? Explain your response.

(b) Suppose we want to find the MLE of β using the Newton-Raphson algorithm. When $\beta^{(t)} = [0, 1]^T$, what is $\beta^{(t+1)}$? Explain your response.

In this exercise, you don't need to show the derivation of the formulas for the gradient and second derivative of the log-likelihood.