Homework 4

ESE 402/542

Due November 17, 2021 at 11:59pm

Type or scan your answers as a single PDF file and submit on Canvas.

Problem 1. Suppose we are given n data points $\{(x_i, y_i)\}_{i=1}^n$, which we assume to be generated

$$y_i = \beta_0 + \beta_1 x_i + \epsilon$$

where $\epsilon \sim \text{dist}(0, \sigma^2)$. Now suppose the data was fit using linear regression by minimizing RSS (least squares), resulting in the two estimated parameters $\hat{\beta}_0, \hat{\beta}_1$. We now want to estimate the y-variable at a new point, x_0. Denoting its true value on the line by $\mu_0 = \beta_0 + \beta_1 x_0$, the estimate is:

$$\hat{\mu}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

(a) Find variance of $\hat{\mu}_0$.

(b) The standard deviation of $\hat{\mu}_0$ can be expressed as a function of $(x_0 - \bar{x})$. Find this function and briefly explain its shape.

(c) Find 95% confidence interval for $\mu_0 = \beta_0 + \beta_1 x_0$ under assumption of normality. (n is large enough)

Problem 2. Assume that $X \sim N(0,1)$, $E \sim N(0,1)$, X and E are independent, and $Y = X + \beta E$. Show that:

$$r_{XY} = \frac{1}{\sqrt{\beta^2 + 1}}$$

Note that r_{XY} is defined as $r_{XY} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$.

Problem 3. Suppose there are n data points $\{x_i \in \mathbb{R}, y_i \in \mathbb{R}\}_{i=1}^n$. We fit a line $y = a + bx$ with least squares, and another line $x = c + dy$ with least squares. Show that $bd \leq 1$, and briefly explain when $bd = 1$ and what it means.
Hint: Cauchy-Schwarz Inequality. $|\text{Cov}(X, Y)|^2 \leq \text{Var}(X) \cdot \text{Var}(Y)$.

Problem 4. (Extra Credit) A student wants to predict a variable, $Y \in \mathbb{R}^n$, from two other variables, $X_1 \in \mathbb{R}^n$ and $X_2 \in \mathbb{R}^n$, using multiple regression. The student defines a new variable $X_3 = X_1 + X_2$ and uses multiple regression to predict Y from X_1, X_2, X_3. Show why this method is problematic.
Hint 1: $A_{n \times n}$ is invertible $\Leftrightarrow$ $\text{Rank}(A) = n$.
Hint 2: $\text{Rank}(AB) \leq \min(\text{Rank}(A), \text{Rank}(B))$.

Problem 5. See the Jupyter notebook file for problem 5.