ECE 792-41 Homework 2
Material Covered: Linear Regression, Normal Equation, Geometric Interpretation

Problem 1 (Least-Squares and Normal Equations) Complete Scheffe 1.1. For ω, you need to do partial differentiation; for Ω, skip the derivation and directly write out the estimate for $[\alpha\ \beta\ \gamma]^T$ using the normal equation in matrix-vector form. Do not simplify.

Problem 2 (Variance and Covariance of Estimators) Complete Scheffe 1.2.

Problem 3 (Least-Squares Estimator vs. Naive Estimators) An ECE student John is doing an electronic circuits lab during which he needs to determine the conductance of a resistor using a voltage meter, a current meter, and a DC power source. The voltage meter is connected in parallel with the resistor and the current meter is in series with the resistor. Both meters are analog devices so the readings recorded by John have errors. The power source is tunable and has a range of 1 to 5 V. Each time John will try a uniformly random input voltage level and record the readings of both voltage and current meters. Denote the voltage reading as x_i and the current reading as y_i for the ith measurement. Assume the true conductance $G = 2\ \text{m}\Omega^{-1}$.

(a) Using a linear model $y_i = Gx_i + e_i$, where e_i is normally distributed with zero-mean and standard deviation $\sigma_e = 0.1$ mA, simulate a dataset of $n = 10$ measurements.

(b) Express the linear model in a matrix-vector form. Clearly indicate what are $\underset{\sim}{y}, \mathbf{X}, \underset{\sim}{\beta}$, and $\underset{\sim}{e}$. Directly implement the formula of the least-squares estimator, $\hat{\underset{\sim}{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\underset{\sim}{y}$, into a computer function that takes as two input vectors $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$, and output a number $\hat{G}$. Apply your function to the simulated data. What is the value of $\hat{G}$? (Hint: $\mathbf{X}$ is a N-by-1 "matrix," and β is a 1-by-1 "vector.")

(c) John's friend proposed a more intuitive estimator for the conductance: $\tilde{G} = \frac{1}{n}\sum_{i=1}^{N}\frac{y_i}{x_i}$. Write a computer function that takes as two input vectors $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$, and output a number $\tilde{G}$. Apply your function to the simulated data. What is the value of $\tilde{G}$?

(d) Find the analytic expression for $\hat{G}$. Show that both $\hat{G}$ and $\tilde{G}$ are unbiased. (Consider y_i as a random variable and x_i as a fixed value.)

(e) Prove that $\text{Var}(\hat{G}) = \left(\sum_{i=1}^{n} x_i^2\right)^{-1}\sigma_e^2$ and $\text{Var}(\tilde{G}) = \left(n^{-2}\sum_{i=1}^{n} x_i^{-2}\right)\sigma_e^2$.

(f) Generate 1,000 datasets. Repeatly apply function written in (b) and collect 1,000 estimates and plot the histogram with 20 bins.

(g) Use the 1,000 datasets generated in (f). Repeatly apply function written in (c) and collect 1,000 estimates and plot the histogram. Which histogram has a narrower spread? Which estimator is better? Why?

Problem 4 (Orthogonal Projection) Consider the set of inconsistent linear equations $\mathbf{A}\underset{\sim}{x} = \underset{\sim}{b}$ given by

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}. \tag{1}$$

(a) Find the least-squares solution to these equations.

(b) Find the "hat" matrix $\mathbf{H}$. Using computer, numerically verify $\mathbf{H} = \mathbf{H}^2$. Argue why.

(c) Find the best approximation $\hat{\underset{\sim}{b}} = \mathbf{H}\underset{\sim}{b}$ to $\underset{\sim}{b}$. Find the vector $\underset{\sim}{b}' = (\mathbf{I} - \mathbf{H})\underset{\sim}{b}$ and show that it is orthogonal to $\hat{\underset{\sim}{b}}$.

(d) What does the matrix $\mathbf{I} - \mathbf{H}$ represent? If $\mathbf{H}$ is called the "orthogonal projector," can you think of a name for $\mathbf{I} - \mathbf{H}$? Numerically verify $\mathbf{I} - \mathbf{H} = (\mathbf{I} - \mathbf{H})^2$. Argue why.

(e) In a 3-dimensional coordinate system, draw the column vectors of matrix $\mathbf{A}$, the column space of $\mathbf{A}$, $\underset{\sim}{b}$, $\hat{\underset{\sim}{b}}$, and $\underset{\sim}{b}'$. Make sure the drawing is reasonably accurate which can reflect the relationship among these quantities.

Problem 5 (Sinusoid as a Predictor) Consider the problem of trying to model a sequence $x(n)$ as the sum of a constant plus a complex exponential of frequency ω_0

$$x(n) \approx c + ae^{jn\omega_0}, \quad n = 0, 1, \ldots, N-1, \tag{2}$$

where c and a are unknown. We may express the problem of finding the values for c and a as one of solving a set of overdetermined linear equations

$$\begin{bmatrix} 1 & 1 \\ 1 & e^{j\omega_0} \\ \vdots & \vdots \\ 1 & e^{j(N-1)\omega_0} \end{bmatrix} \begin{bmatrix} c \\ a \end{bmatrix} = \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix}. \tag{3}$$

(a) Find the least-squares solution for c and a. Note that when dealing with problem with complex numbers, the transpose operator in the normal equation will become the Hermitian operator, i.e., transpose + complex conjugate.

(b) If N is even and $\omega_0 = 2\pi k/N$ for some integer k, find the least-squares solution for c and a.

Problem 6 (Data-Driven Formula Search) It is known that the sum of the squares of n from $n = 0$ to $N - 1$ has a closed-form solution of the following form

$$\sum_{n=0}^{N-1} n^2 = a_0 + a_1 N + a_2 N^2 + a_3 N^3. \tag{4}$$

Given that a third-order polynomial is uniquely determined in terms of the values of the polynomial at four distinct points, derive a closed-form expression for this sum by setting up a set of linear equations and solving these equations for a_0, a_1, a_2, and a_3. Verify your solution against the formula $\sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6$.