ECE 411 Homework 3 (Fall 2024)
Instructor: Dr. Chau-Wai Wong
Material Covered: Linear Algebra, Geometric Interpretation

Problem 1 (20 points) [Implementing Least-Squares Estimator]

a) Dataset Generation. A course has $N = 20$ students. The ith student spends x_{1i} hours on homework, x_{2i} hours on reviewing the course material, and received a score of y_i at the exam. x_{1i} is drawn from a uniform distribution Uniform$(2, 8)$ and x_{2i} is drawn from a uniform distribution Uniform$(3, 16)$. We assume that the true score is generated by the following linear statistical model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i, \quad i = 1, \ldots, N, \tag{1}$$

where e_i is a zero-mean Gaussian random variable with standard deviation σ. Now, use a programming language of your choice, implement a computer function named `DatasetGeneration` with the following input and output signatures:

`Inputs: beta_vector, standard_deviation`,
`Outputs: label_vector_Y, feature_vector_X1, feature_vector_X2`.

Here, `beta_vector` is an array of length 3, and `label_vector_Y`, `feature_vector_X1`, and `feature_vector_X2` are arrays of length 20.

Set input parameters `beta_vector` and `standard_deviation` to some values and call the `DatasetGeneration` function. Report your choice of the input parameters and the returned output values. Attach the source code of the function.

b) Estimator Implementation. Implement `LeastSquaresEstimator` using the matrix–vector form of the normal equation. You do not need to derive the normal equation but please ensure the data matrix $\mathbf{X}$ and label vector $\mathbf{Y}$ in the normal equation are correctly represented in your computer code. The function's inputs and outputs are as follows:

`Inputs: label_vector_Y, feature_vector_X1, feature_vector_X2`,
`Outputs: beta_hat_vector`.

Use the dataset generated in part a), report the estimated parameters $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$. Attach the source code of the function.

c) Quality Evaluation. Fix `beta_vector` and `standard_deviation`. Repeatedly run `DatasetGeneration` followed by `LeastSquaresEstimator` for 1,000 times. Store the estimated parameters in arrays. Use the built-in functions, calculate the sample mean and sample standard deviation for each estimated parameter. Attach the source code. Compare the sample mean of each estimated parameter to the value of the true parameter you specified. Did the least-squares estimator do a good job?

Problem 2 (20 points) [Linear Independence, Basis, and Vector Space]

a) Are vectors $[9\ \ 8], [7\ \ 6]$, and $[5\ \ 4]$ linearly independent? What about $[-1\ \ 1\ \ 0], [1\ \ -1\ \ 1]$, and $[0\ \ 0\ \ 2]$? Justify your answers.

b) You are given a vector space $V = \text{span}\,\{[-1\ \ 0\ \ 0], [0\ \ -1\ \ 0], [1\ \ 1\ \ 0]\}$.

(i) Express V in a set representation.

(ii) Can you find a basis for V?

(iii) Are $[4\ \ 1\ \ 0]$, $[1\ \ 0\ \ 4]$, and $[0\ \ 4\ \ 1]$ in vector space V? Is yes, what are the coefficient for each vector of the basis you found in (ii)?

(iv) Draw all points of (iii) in a 3D coordinate. Illustrate vector space V using a plane formed by the vectors of the basis.

c) Let

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}.$$

What is the dimension of the row vector space of $\mathbf{A}$? What is the row rank of $\mathbf{A}$?

Problem 3 (10 points, Bonus) [Linear Independence and Approximate Orthogonality of High-Dimensional Random Vectors]

a) (5 points, Bonus) Find a physical coin and you will use it to generate binary values: -1 (tail) or 1 (head). Toss the coin for 30 times and record the outcomes. Put the outcomes into 6 vectors of length 5. (Make sure no two vectors are exactly the same. If yes, discard one vector and toss the coin to generate a new vector.) Verify that the first 5 vectors are linearly independent. Verify that the 6 vectors are linearly dependent. (You may use a computer to help the verification.) Could you explain why adding the 6th vector makes the set of vectors linearly dependent?

b) (5 points, Bonus) Use your favorite programming language to generate 100 vectors of length 100. Each vector is generated as follows: (i) each entry of the vector is a random number drawn from a Gaussian distribution with zero mean and unit variance, and (ii) the vector is then scaled to have a unit norm. Calculate the pairwise inner product between any two vectors. You should obtain 4,950 inner product values. Plot the histogram of these inner product values. You should see that the histogram is centered around zero and has a narrow peak.

Problem 4 (20 points) [Softmax Function] Given an input image, a neural network extracts a sequence of features $\mathbf{z} = (z_1, \ldots, z_K)$. The softmax output for the ith feature z_i is given by

$$\sigma_i(\mathbf{z}) = \frac{\exp(\beta z_i)}{\sum_{j=1}^{K} \exp(\beta z_j)},$$

where β is a positive integer.

a) When $\mathbf{z} = (1, 2, 3, 4, 5)$, use your favorite programming language to calculate and plot $\sigma_i(\mathbf{z})$ as a function of i in bar charts when β takes values of 0.1, 1, and 10, respectively. Based on the empirical results, could you guess what the role of β is?

b) Prove that $\left(\sigma_1(\mathbf{z}), \ldots, \sigma_K(\mathbf{z})\right)$ is a valid probability mass function.

c) When z_1 is the largest feature value, prove that $\sigma_1(\mathbf{z}) = 1$ as $\beta \to \infty$.

d) When z_1 is the largest feature value, prove that $\sigma_j(\mathbf{z}) = 0$ for $j = 2, \ldots, K$ as $\beta \to \infty$.

e) Show that $\sigma_j(\mathbf{z}) = 1/K$ for all $j \in [1, K]$ when $\beta = 0$.

f) How are the results in c)–e) connected to your guess about the role of β in a)?