ECE 492-45 Homework 3 (Fall 2021)
Instructor: Dr. Chau-Wai Wong
Material Covered: Geometric interpretation, Modern ML applications

Problem 1 (20 points) [Orthogonal Projection] Consider the set of inconsistent linear equations $\mathbf{Ax} = \mathbf{b}$ given by

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

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

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

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

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 vector space of $\mathbf{A}$, $\mathbf{b}$, $\widehat{\mathbf{b}}$, and $\mathbf{b}'$. Make sure that the drawing is reasonably accurate which can reflect the relationship among these quantities.

Problem 2 (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 $\Big(\sigma_1(\mathbf{z}), \ldots, \sigma_K(\mathbf{z})\Big)$ 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)?

This homework has only two problems.