CSE 6363: Machine Learning
University of Texas at Arlington
Fall 2025
Alex Dillhoff

Problem Set 1

These problems are meant to fill any mathematical gaps you may have and to help familiarize you to the notation and concepts we will be using in this course. You may use online resources, books, or any other material to help you solve these problems. You may also discuss the problems with your classmates, but you must write up your solutions on your own.

1. Use maximum likelihood estimation to derive the mean and variance of a Gaussian distribution.

2. **(Variant of Bishop 1.1)** Consider a linear model give by

$$y(x, \mathbf{w}) = w_0 + w_1 x,$$

where w_0 and W_1 are the parameters of the model, and a training set of N data points $\{x_n, t_n\}$. The sum-of-squares function is defined as

$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - y(x_n, \mathbf{w})\}^2.$$

Show that the coefficients $\mathbf{w} = (w_0, w_1)$ that minimize this error function are given by the solution to the following set of linear equations

$$\sum_{j=0}^{1} A_{ij} w_j = T_i$$

where

$$A_{ij} = \sum_{n=1}^{N}(x_n)^{i+j}$$
$$T_i = \sum_{n=1}^{N}(x_n)^{i} t_n.$$

Hint: These equations represent the normal equations for the least squares solution.

3. Show that the derivative of the cross-entropy function is the same as the derivative of the least squares error function.

4. Derive the likelihood function for a Categorical distribution using the probability mass function of a multinomial distribution.

5. Prove that the inverse of the sigmoid function, $\sigma(x) = \frac{1}{1+e^{-x}}$, is the logit function, $\text{logit}(x) = \log\left(\frac{x}{1-x}\right)$.

You may submit your work as either a scanned PDF OR you may take pictures of your solutions and combine them into a PDF. **Do not submit individual images. Rename your submission as LASTNAME_ID_A1.pdf**. For example, Naomi Nagata (7355608) would save their file as `NAGATA_7355608_A1.pdf`.