Problem 1

Find the derivative of the sigmoid function with respect to \( x \) where the sigmoid function \( \sigma(x) \) is given by,

\[
\sigma(x) = \frac{1}{1 + e^{-x}}
\]

Problem 2

1. When are two vectors \( u \) and \( v \in \mathbb{R}^n \) said to be orthogonal?
2. Are the following vectors orthogonal to each other?

\[
v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad v_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad v_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}
\]

Problem 3

Consider the following Bayesian Network containing four Boolean random variables.

![Bayesian Network Diagram]

1. Compute \( P(A|C) \)
2. Compute \( P(\neg A, B, \neg C, D) \)

Problem 4

In order to get an unbiased estimate of how well your ML algorithm is doing. A typical split is 60/20/20 split in terms of 60% train, 20% test and 20% validation. And several years ago this was widely considered best practice in machine learning. Do you still agree such ratios in the modern big data era? Why or why not?
Problem 5

Suppose you’re training a neural network in an unusual, nondeterministic domain: The training set consists of \( N \) copies of the same example, a fraction \( p > 0.5 \) of which are positive and a fraction \( 1 - p \) of which are negative. Suppose we decide we want to optimize the absolute error function

\[
E = \sum_{i=1}^{N} |T_i - O|
\]

where \( T_i \) is the correct value for example \( i \) and \( O \) is the network’s output for this example. Suppose that \( O \) must also be in the range \([0,1]\). By writing out an expression for the error in terms of \( O \), find the value of \( O \) that minimizes the error.