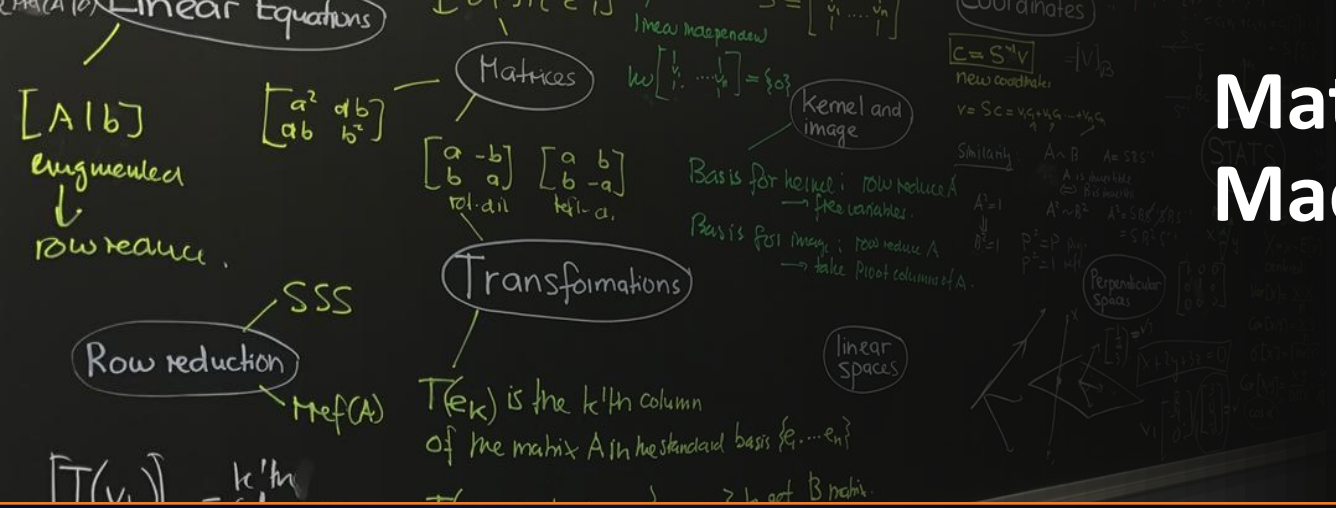



# Mathematical Foundations for Machine Learning and Data Science

## Tutorial 06


Dr. Zubair Khalid

Department of Electrical Engineering
School of Science and Engineering
Lahore University of Management Sciences

https://www.zubairkhalid.org/ee212_2021.html

LUMS
A Not-for-Profit University

# Problem 01:

Consider a network of $N$ links, labeled $1, 2, \ldots, n$. A path through the network is a subset of the links. (The order of the links on a path does not matter here.) Each link has a (positive) delay, which is the time it takes to traverse it. We let $d$ denote the $n$-vector that gives the link delays. The total travel time of a path is the sum of the delays of the links on the path. Our goal is to estimate the link delays (i.e., the vector $d$), from a large number of (noisy) measurements of the travel times along different paths. This data is given to you as an $N \times n$ matrix $P$, where

$$P_{ij} = \begin{cases} 1 & \text{link } j \text{ is on path } i \\ 0 & \text{otherwise} \end{cases}$$

and an $N$-vector $t$ whose entries are the (noisy) travel times along the $N$ paths. You can assume that $N > n$. You will choose your estimate $\hat{d}$ by minimizing the RMS deviation between the measured travel times ($t$) and the travel times predicted by the sum of the link delays.

(a) Formulate the problem of determining $\hat{d}$ as least-squares problem.

(b) Provide a matrix expression for $\hat{d}$, that is, solve the least-squares problem and express $\hat{d}$ in terms of $P$ and $t$.

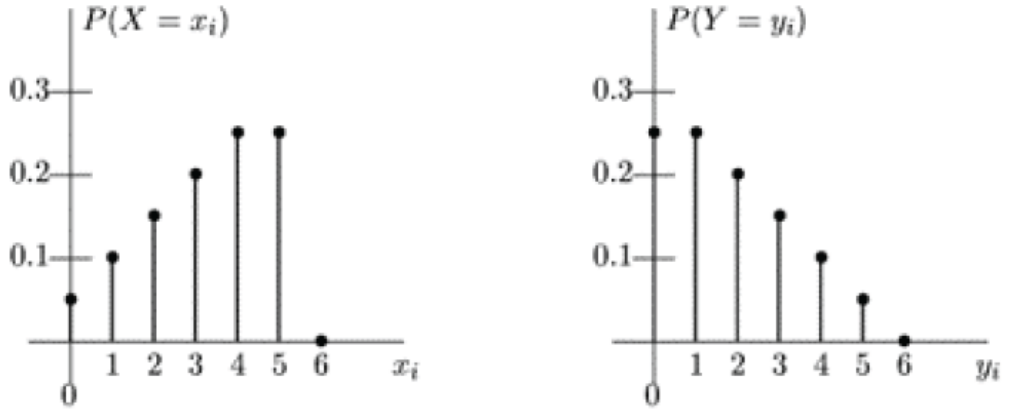
## Problem 02:

Three different models are fitted using the same training data set, and tested on the same (separate from the training set) test set (which has the same size as the training set). The RMS prediction errors for each model, on the training and test sets, are reported below. Comment briefly on the results for each model. You might mention whether the model's predictions are good or bad, whether it is likely to generalize to unseen data, or whether it is over-fit.

| Model | Train RMS | Test RMS |
|---|---|---|
| A | 0.135 | 2.125 |
| B | 0.680 | 0.675 |
| C | 30.03 | 1.392 |

## Problem 03:

Let $X$ be a RV that denotes the number of faulty products in a shipment shipped from Apple, and let $Y$ be the number of faulty products in a shipment shipped from Samsung. The two companies have known probability mass functions which give the probability of a given number of faulty products in a given order. They display their information graphically, as follows:




(a) Suppose you receive a shipment of products from Apple. Find the probability that the number of faulty products is less than or equal to 3.

(b) Suppose you receive a shipment of products from Apple. Somebody tells you that there are at least two faulty products in the shipment. Given this information, find the probability that the number of faulty products is less than or equal to 3.

(c) Suppose a shipment of products arrives, but you are not sure from which company it came; however, you know the probability it came from Apple is 0.35 and the probability it came from Samsung is 0.65. Find the probability that the number of faulty products in the shipment is less than or equal to 3.

## Problem 04:

Consider a function $f : \mathbf{R}^n \to \mathbf{R}$ given by

$$f(x) = x^T P x, \quad x \in \mathbf{R}^n,\ P \in \mathbf{R}^{n \times n}$$

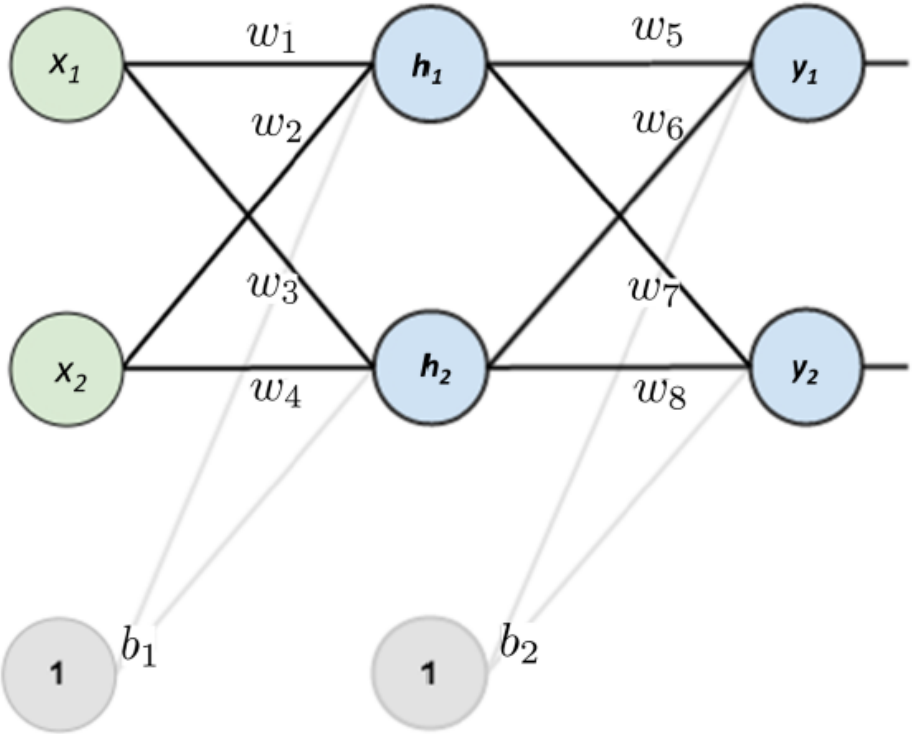
(a) Determine the gradient of the function $f$.

(b) Determine the Hessian matrix of the function $f$.

(c) What are the conditions on $P$ for a function $f$ to be convex?

# Problem 05:

Consider a two-layer neural network with with two inputs, two outputs and two neurons in the hidden layer with interconnections shown in the figure below. We assume that the sigmoid function has been used as the activation function.




(a) Express the output $y_1$ in terms of inputs $x_1$ and $x_2$. You may write a series of equations that relate the output $y_1$ to the inputs.