EML Assignment 4 Problem 1

Muhammed Saeed and Ali Salaheldin Ali Ahmed

December 2022

Question 1

minimising the projection residuals, we can specify the line by a unit vector $\vec{w}$ and then the projection of data vector $barx$ on the line is $\cdot(\vec{x}.\vec{w})$. which is the distance between the projection to the origin, the actual coordinate in p-space is $(\vec{x} \cdot \vec{x})\vec{w}$. the mean of the projections should be zero since the vectors are scaled and centered around the zero

$$\frac{\sum_{i=1}^{n}(\vec{x} \cdot \vec{w})\vec{w})}{n} \tag{1}$$

The error related to the projection is given by
$\|\vec{x_i} - (\overrightarrow{x + i} \cdot \vec{w})\vec{w}\| = \vec{x_i} \cdot \vec{x_i} - (\vec{w}\vec{x_i})^2\|$
$MSE = \frac{(\sum_{i=1}^{n}\|\vec{x_i}\|^2 - \sum_{i=1}^{n}(\vec{w}\vec{x_i})^2}{n}$

The first term does not depend on w therefore in order to minimize the MSE we need to make the second term larger, we also know that the mean of the square is the sqaure of the mean plus the variance
$\sum_{i=1}^{n}((\vec{w}\vec{x_i})^2)/n = (\frac{\sum_{i=1}^{n}(\vec{w}\vec{x_i})}{n})^2 + Var[\vec{w} \cdot \vec{x}]$
Since we've just seen that the mean of the projections -first term - is zero, minimizing the residual sum of squares turns out to be equivalent to maximizing the variance of the projections.