COMP527
Data Mining and Visualisation
Problem Set 3

Danushka Bollegala

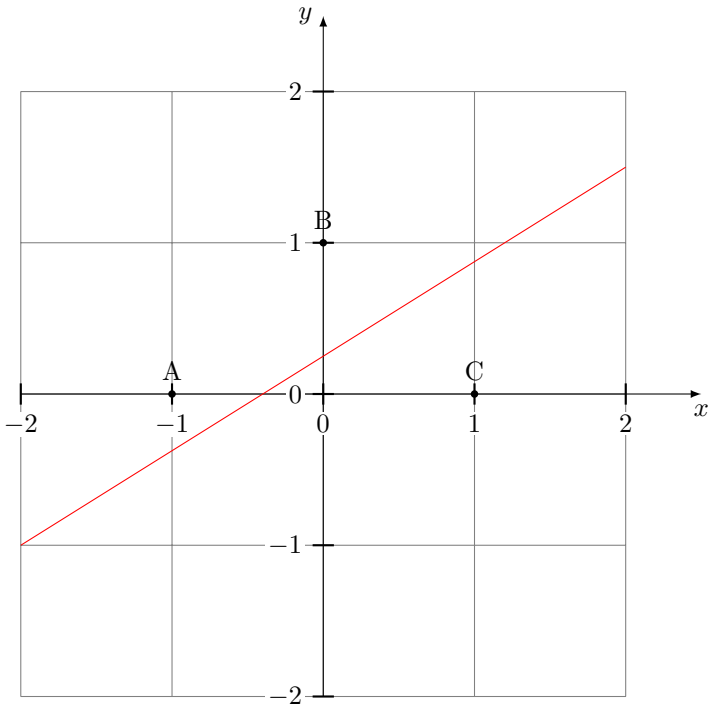

Figure 1: Projecting three points A, B, C onto the line $y = mx + c$.

Question Consider the problem of projecting a two-dimensional dataset consisting of three points $A = (-1, 0), B = (0, 1)$, and $C = (1, 0)$ onto the one-dimensional line given by $y = mx + c$. The dataset and the line is shown in Figure 1. Answer the following questions.

A. Compute the co-ordinates of the base of the perpendicular from point (α, β) to line $y = mx + c$. **(10 marks)**

B. Compute the perpendicular distance to the line $y = mx + c$ from point (α, β). **(10 marks)**

C. Show that if $y = mx + c$ is a solution to the one dimensional PCA projection, then $y = mx + c'$ is also a solution. Here, $c \neq c'$. **(10 marks)**

D. Find m such that the variance of the projected points on to the straight line is maximised. **(20 marks)**

E. Find m such that the sum of squared projection errors is minimised. **(20 marks)**

F. Compute the covariance matrix for this dataset. **(10 marks)**

G. Find the eigenvalues and eigenvectors of the covariance matrix. **(10 marks)**

H. Find the PCA projection using the eigenvalue decomposition of the covariance matrix. **(10 marks)**