1 Part 1

1. Consider the following supervised dataset consisting of 10 observations, each with two features (observable data is in X) and an associated label in Y:

$$\text{X} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 1 \\ 0 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 2 & 0 \\ 2 & 1 \end{bmatrix} \text{ Y} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

   (a) First Feature = 0.4854752972273343

$$0\text{: } y = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \tfrac{3}{10}(-\tfrac{3}{3}log_2\tfrac{3}{3} - \tfrac{0}{3}log_2\tfrac{0}{3}) = 0$$

$$1\text{: } y = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tfrac{5}{10}(-\tfrac{2}{5}log_2\tfrac{2}{5} - \tfrac{3}{5}log_2\tfrac{3}{5}) = 0.4854752972273343$$

$$2\text{: } y = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tfrac{2}{10}(-\tfrac{0}{2}log_2\tfrac{0}{2} - \tfrac{2}{2}log_2\tfrac{2}{2}) = 0$$

   (b) Second Feature = 0.9709505944546686

$$0\text{: } y = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tfrac{5}{10}(-\tfrac{2}{5}log_2\tfrac{3}{5} - \tfrac{3}{5}log_2\tfrac{3}{5}) = 0.4854752972273343$$

$$1\text{: } y = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \tfrac{5}{10}(-\tfrac{3}{5}log_2\tfrac{3}{5} - \tfrac{2}{5}log_2\tfrac{3}{5}) = 0.4854752972273343$$

   (c) Feature 1 had a lower score so it was better at class seperation.

   (d) I found each feature's corresponding mean and standard deviation. Then I standardized them accordingly to get the new standardized matrix:

$$X = \begin{bmatrix} -1.21973567 & 0.0513167 \\ -1.21973567 & -0.9486833 \\ -0.21973567 & 0.0513167 \\ -1.21973567 & -0.9486833 \\ -0.21973567 & 0.0513167 \\ -0.21973567 & -0.9486833 \\ -0.21973567 & -0.9486833 \\ -0.21973567 & 0.0513167 \\ 0.78026433 & -0.9486833 \\ 0.78026433 & 0.0513167 \end{bmatrix}$$

After that I then calculated the covariance matrix to get the corresponding eigen values and vectors. I used the formula $Z = XW$ to get the pca where the first column is the first feature projection and so on:

$$pca = \begin{bmatrix} -1.18598519 & 0.28953001 \\ -1.38210133 & -0.69105066 \\ -0.20540452 & 0.09341388 \\ -1.38210133 & -0.69105066 \\ -0.20540452 & 0.09341388 \\ -0.40152065 & -0.8871668 \\ -0.40152065 & -0.8871668 \\ -0.20540452 & 0.09341388 \\ 0.57906002 & -1.08328293 \\ 0.77517616 & -0.10270226 \end{bmatrix}$$

(e) It currently is still within the 2D coordinate system because we didn't reduce our dimensionality from the original.

(f) The data matrix would just be the first single column of the pca in order to reduce it to 1D.

$$1D = \begin{bmatrix} -1.18598519 \\ -1.38210133 \\ -0.20540452 \\ -1.38210133 \\ -0.20540452 \\ -0.40152065 \\ -0.40152065 \\ -0.20540452 \\ 0.57906002 \\ 0.77517616 \end{bmatrix}$$

2 Part 2