1. Theory Questions

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

$$\mathbb{E}(\mathrm{H(A)}) = \sum_{i=1}^{k} \frac{p_i + n_i}{p+n} * \boldsymbol{H}\left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i}\right)$$

$$H\big(P(v_1), P(v_2)\big) = \sum\nolimits_i^n (-P(v_i)\log_2 P(v_i))$$

$$\mathrm{C1} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}, \mathrm{C0} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 2 & 0 \\ 2 & 1 \end{bmatrix}$$

a) Entropy Based On Feature 1:

N = 5, P = 5

When $f_1 = 0$

$$\mathrm{X} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \; Y = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$P_0 = 3$, $N_0 = 0$, P(Y = 1 | f_1 = 0) = 1 , P(Y = 0 | f_1 = 0) = 0

When $f_1 = 1$

$$\mathrm{X} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \; Y = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$P_1 = 2$, $N_1 = 3$, P(Y = 1 | f_1 = 1) = 2/5 , P(Y = 0 | f_1 = 1) = 3/5

When $f_1 = 2$

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

$P_2 = 0$, $N_2 = 2$, P(Y = 1 | f_1 = 2) = 0 , P(Y = 0 | f_1 = 1) = 1

$$\mathbb{E}(\mathrm{H}(f_1)) = \left(\frac{3+0}{5+5} * (-1 * \log_2 1 - \ 0 * \log_2 0\ )\right) + \left(\frac{2+3}{5+5} * (-\frac{2}{5} * \log_2 \frac{2}{5} - \frac{3}{5} * \log_2 \frac{3}{5})\right) + \left(\frac{0+2}{5+5} * (0 * \log_2 0 - \ 1 * \log_2 1\ )\right) = \ .48547$$

b) Entropy Based On Feature 2:

N = 5, P = 5

When $f_2 = 0$

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

$P_0 = 2$, $N_0 = 3$, $P(Y = 1 \mid f_2 = 0) = 2/5$, $P(Y = 0 \mid f_2 = 0) = 3/5$

When $f_2 = 1$

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

$P_1 = 3$, $N_1 = 2$, $P(Y = 1 \mid f_2 = 1) = 3/5$, $P(Y = 0 \mid f_2 = 1) = 2/5$

$$\mathbb{E}(\mathrm{H}(f_2)) = \left(\frac{2+3}{5+5} * (-\frac{2}{5} * \log_2 \frac{2}{5} - \frac{3}{5} * \log_2 \frac{3}{5})\right) + \left(\frac{3+2}{5+5} * (-\frac{3}{5} * \log_2 \frac{3}{5} - \frac{2}{5} * \log_2 \frac{2}{5})\right) = \ .97095$$

c) Based on a and b, which feature is more discriminating

The more discriminating feature is the feature with the lower entropy, which means feature 1 (f1) is more discriminating than feature 2 (f2)

d) What are the principal components of the observed data X?

$$Z\ Scored\ X = \begin{bmatrix} -1.21973567 & 0.9486833 \\ -1.21973567 & -0.94868330 \\ 0.13552619 & 0.9486833 \\ -1.21973567 & -0.94868330 \\ 0.13552619 & 0.9486833 \\ 0.13552619 & -0.9486833 \\ 0.13552619 & -0.94868330 \\ 0.135526191 & 0.94868331 \\ 1.490788042 & -0.9486833 \\ 1.49078804 & 0.9486833 \end{bmatrix}, \quad Y = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\text{Covariance Of X} = \begin{bmatrix} 1 & 0.142857140 \\ 0.14285714 & 1 \end{bmatrix}$$

$$EigenValues = \ 1.14285714, 0.85714286$$

$$EigenVectors = \begin{bmatrix} -0.70710678 \\ -0.70710678 \end{bmatrix}, \begin{bmatrix} -0.70710678 \\ 0.70710678 \end{bmatrix}$$

The two principle components are EigenVectors listed above, $\begin{bmatrix} -0.70710678 \\ -0.70710678 \end{bmatrix}$ and $\begin{bmatrix} -0.70710678 \\ 0.70710678 \end{bmatrix}$, calculated by z scoring the X data, then getting the covariance matrix of X^T, then used a Single Value Decomposition (SVD) function provided by the numpy library to get the Eigenvalues and Eigenvectors listed.

e) Describe these axis in terms of a conventional 2D Cartesian Coordinate System

What these principle components essentially mean is that instead of having a normal 2D Cartesian Coordinate System like we are used to with the X and Y's going in the completely horizontal and vertical directions, we have a coordinate systems where the axis are shifted based on the eigenvectors. So when we look at our new coordinate system, instead of the X axis just going straight horizontal (when look at with the Cartesian System) we instead have that new "X" axis following $\begin{bmatrix} -0.70710678 \\ -0.70710678 \end{bmatrix}$ and the new "Y" axis is following $\begin{bmatrix} -0.70710678 \\ 0.70710678 \end{bmatrix}$. We are basically creating a new plane for our data to sit in which allows us to reduce the dimensionality of it if we so desire.

d) If we were to project our data down to 1-D using the principle component, what would the new data matrix X be

Since we are projecting down to 1-D, we would want to use the eigenvector which corresponds with the higher eigenvalue, so in this case the larger eigenvalue of 1.14285714 corresponds with the eigenvector of $\begin{bmatrix} -0.70710678 \\ -0.70710678 \end{bmatrix}$, thus, to get out our new X we would multiple X by that eigenvector

$$New\ X = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 1 \\ 0 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 2 & 0 \\ 2 & 1 \end{bmatrix} * \begin{bmatrix} -0.70710678 \\ -0.70710678 \end{bmatrix} = \begin{bmatrix} -0.707106780 \\ 0 \\ -1.41421356 \\ 0 \\ -1.41421356 \\ -0.70710678 \\ -0.70710678 \\ -1.41421356 \\ -1.41421356 \\ -2.12132034 \end{bmatrix}$$

2. Faces Graph

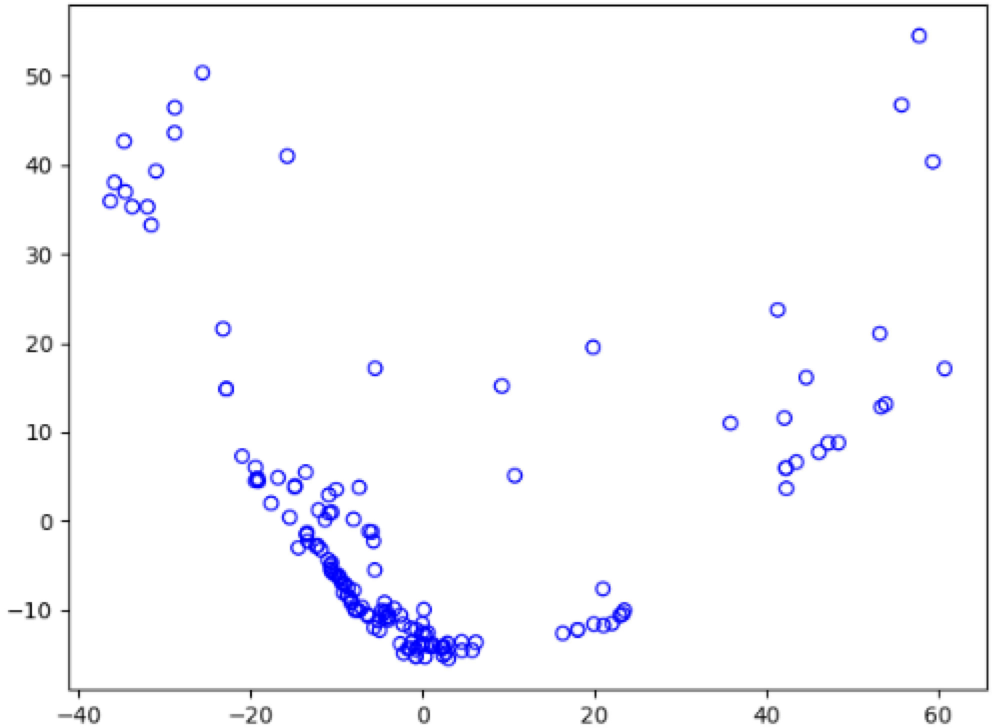

Figure 1: Result of Plotting Yales Faces In 2D Using PCA