# Dimensionality Reduction

Ahnaf An Nafee

February 2021

## 1 Theory Questions

1. Datasets:

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

   (a) Computing the average weighted entropy (feature 1):

$p_0 = 3 \quad n_0 = 0$
$p_1 = 2 \quad n_1 = 3$
$p_2 = 0 \quad n_2 = 2$

$$\begin{aligned} E(H(1)) = \frac{3+0}{10} \times (\frac{-3}{3} \cdot log_2 \frac{3}{3} - 0+ \\ \frac{2+3}{10} \times (\frac{-2}{5} \cdot log_2 \frac{2}{5} + \frac{-3}{5} \cdot log_2 \frac{3}{5})+ \\ \frac{0+2}{10} \times (0 + \frac{-2}{2} \cdot log_2 \frac{2}{2}) \end{aligned} \tag{1}$$

$$E(H(1)) = 0.4855 \tag{2}$$

(b) Computing the average weighted entropy (feature 2):

$$\begin{array}{ll} p_0 = 2 & n_0 = 3 \\ p_1 = 3 & n_1 = 2 \\ p_2 = 0 & n_2 = 0 \end{array}$$

$$\begin{aligned} E(H(2)) = \frac{2+3}{10} \times (\frac{-2}{5} \cdot log_2\frac{2}{5} + \frac{-3}{5} \cdot log_2\frac{3}{5}) + \\ \frac{3+2}{10} \times (\frac{-3}{5} \cdot log_2\frac{3}{5} + \frac{-2}{5} \cdot log_2\frac{3}{5}) + \\ 0 \times (0+0) \end{aligned} \tag{3}$$

$$E(H(2)) = 0.9710 \tag{4}$$

(c) Feature 1 is more discriminating

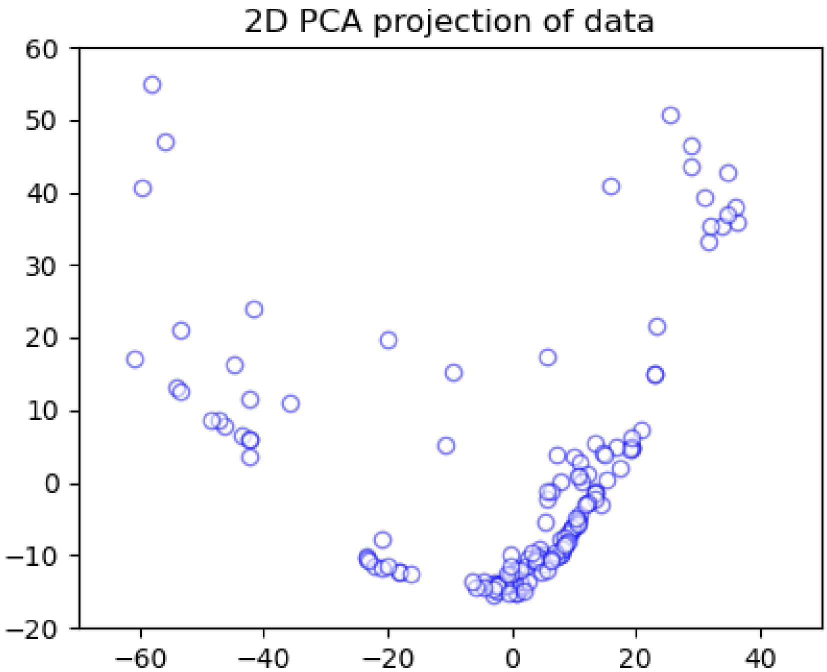
(d) Principal Components:

```
[[ 0.70710678 -0.70710678]
 [ 0.70710678  0.70710678]]
```

(e) The x-axis corresponds to the first principal component while the y-axis corresponds to the second principal component

(f) PCA (1D) = $\begin{bmatrix} -0.19166297 \\ -1.53330376 \\ 0.76665188 \\ -1.53330376 \\ 0.76665188 \\ -0.57498891 \\ -0.57498891 \\ 0.76665188 \\ 0.38332594 \\ 1.72496673 \end{bmatrix}$

## 2 Dimensionality Reduction via PCA




## 3 Eigenfaces

1. `See code in the Jupyter Notebook`