1. Theory

a. average weighted entropy of first col

$$g = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad y = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad g = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$H_0 = 0$
$H_2 = 0$

$E = \frac{3}{10}(0) + \frac{5}{10}(0.97) + \frac{2}{6}(0)$

$E = 0.485$

$H_1 = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log \frac{2}{5}$

b. f_2

$$y = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad y = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

$E = \frac{5}{10}(0.97) + \frac{5}{10}(0.97)$

$E = 0.97$

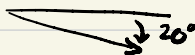$h_0 = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log \frac{2}{5}$

$h_1 = -\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5}$

c. The first feature is more descriminating because it has a lower entropy and offers better separation.

d. the principal component in unit length are:

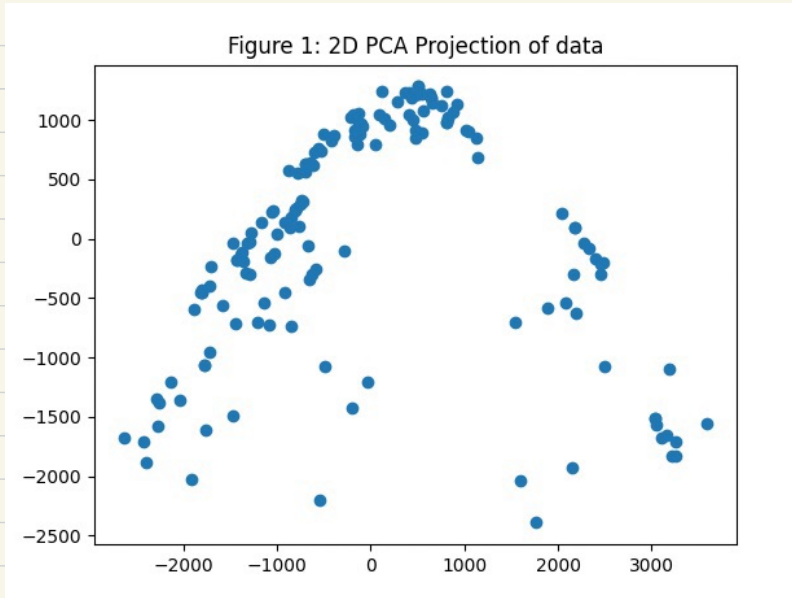$$\begin{bmatrix} 0.98 & -0.19 \\ 0.19 & 0.98 \end{bmatrix}$$

e. it looks like the horizontal axis has been tilted 20° down.

→ 20°

f.
$$x = \begin{bmatrix} -7.84e-1 \\ -9.805e-1 \\ 1.96e-1 \\ -9.80e-1 \\ 1.96e-1 \\ -2.17e-17 \\ -2.17e-17 \\ 1.96e-1 \\ 9.8e-1 \\ 1.176 \end{bmatrix}$$

> I assume these two are just zero.

2.

Figure 1: 2D PCA Projection of data

3. code only