Alec Peterson
ap3842@drexel.edu

CS 613 – Homework 1

1) **Problem 1**

a. Weight-averaged entropy of feature 1:

$$P_{x=0,y=0} = \frac{0}{3} = 0\,, P_{x=0,y=1} = \frac{3}{3} = 1$$
$$P_{x=1,y=0} = \frac{3}{5}\,, P_{x=1,y=1} = \frac{2}{5}$$
$$P_{x=2,y=0} = \frac{2}{2} = 1\,, P_{x=2,y=1} = \frac{0}{2} = 0$$

$$H_1 = -(0)(\log_2(0)) - (1)(\log_2(1)) = 0$$
$$H_2 = -\left(\frac{3}{5}\right)\log_2\left(\frac{3}{5}\right) - \left(\frac{2}{5}\right)\log_2\left(\frac{2}{5}\right) = 0.971$$
$$H_3 = -(1)(\log_2(1)) - (0)(\log_2(0)) = 0$$

$$E_{feature\ 1} = \sum_{i=1}^{3}\frac{|C_i|}{N}H_i = \frac{3}{10}(0) + \frac{5}{10}(0.971) + \frac{2}{10}(0) = 0.515$$

b. Weight-averaged entropy of feature 2:

$$P_{x=0,y=0} = \frac{3}{5}\,, P_{x=0,y=1} = \frac{2}{5}$$
$$P_{x=1,y=0} = \frac{2}{5}\,, P_{x=1,y=1} = \frac{3}{5}$$

$$H_1 = -\left(\frac{3}{5}\right)\log_2\left(\frac{3}{5}\right) - \left(\frac{2}{5}\right)\log_2\left(\frac{2}{5}\right) = 0.971$$

$$H_2 = -\left(\frac{2}{5}\right)\log_2\left(\frac{2}{5}\right) - \left(\frac{3}{5}\right)\log_2\left(\frac{3}{5}\right) = 0.971$$

$$E_{feature\ 2} = \sum_{i=1}^{2}\frac{|C_i|}{N}H_i = \frac{5}{10}(0.971) + \frac{5}{10}(0.971) = 0.971$$

c. Feature 1 provides a more discriminating result i.e. better class separation as it gave a weight-averaged normalized entropy of approximately 0.51, lower than that of (and indicating it is less random than) Feature 2 which gave a weight-averaged entropy of approximately 0.97.

d. Principal components of observed data X, found after mean-centering data, calculating covariance matrix of the mean-centered data, and then calculating the eigenvectors using numpy.linalg.eig() and ordered based on decreasing eigenvalue:

(rounded values)

1^{st} Principal Component: [0.981, 0.196]
2^{nd} Principal Component: [-0.196, 0.981]

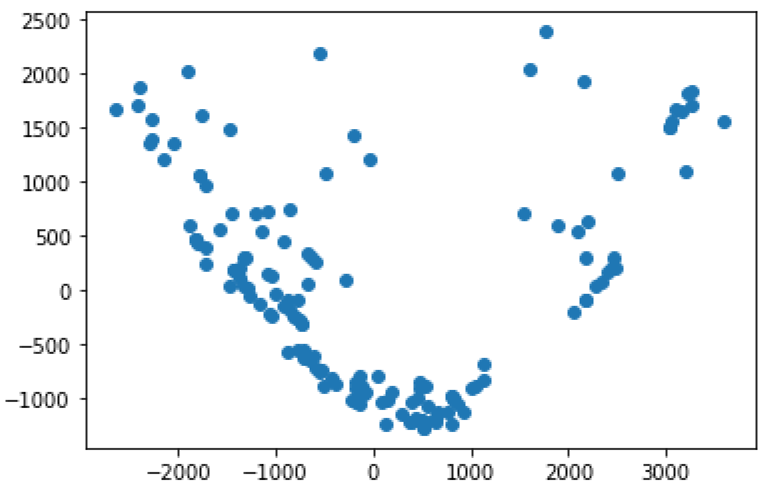
e. The "axes" formed by the principal components (orthogonal eigenvectors of the covariance matrix) form a transformed 2D coordinate system i.e. basis onto which points from the original matrix X could be projected. These axes point in the directions of most variance and in being orthogonal, reduce collinearity present in the original coordinate system.

f. Projected data matrix onto 1^{st} principal component (rounded values):

[-0.78, -0.98, 0.20, -0.98, 0.20, 0.00, 0.00, 0.20, 0.98, 1.18]

2) Problem 2

Results from reducing data to 2D using PCA and plotting via matplotlib.pyplot.scatter() – see HW1_Problem2.py



3) **Problem 3 –** see submitted video and HW1_Problem3.py