CS 383 - Machine Learning

Assignment 1 - Dimensionality Reduction

1 (25pts) Theory Questions

$$X = \begin{bmatrix} -2 & 1 \\ -5 & -4 \\ -3 & 1 \\ 0 & 3 \\ -8 & 11 \\ -2 & 5 \\ 1 & 0 \\ 5 & -1 \\ -1 & -3 \\ 6 & 1 \end{bmatrix} \quad Y = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

(a) Compute the average entropy of the class label in the subsets created by splitting the dataset based on the value of the first feature. You may assume that the features are *categoricial*. (5pts).

$p_{-2} = 1 \qquad n_{-2} = 1$
$p_{-5} = 1 \qquad n_{-5} = 0$
$p_{-3} = 1 \qquad n_{-3} = 0$
$p_0 = 1 \qquad n_0 = 0$
$p_{-8} = 1 \qquad n_{-8} = 0$
$p_{-2} = 1 \qquad n_{-2} = 1$
$p_1 = 0 \qquad n_1 = 1$
$p_5 = 0 \qquad n_5 = 1$
$p_{-1} = 0 \qquad n_{-1} = 1$
$p_6 = 0 \qquad n_6 = 1$

$$E(H(A)) \sum_{i=1}^{K} \frac{p_i + n_i}{p + n} H\left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i}\right)$$

Feature 1 (first index):

$$E(H(1)) \sum_{i=1}^{1} \frac{1+1}{5+5} * \left(-\frac{1}{1+1} log_2 \left(\frac{1}{2}\right) + \frac{-1}{1+1} log_2 \left(\frac{1}{2}\right)\right) = 0.2$$

Feature 1 (second index):

$$E(H(1)) \sum_{i=2}^{2} \frac{1}{10} * \left(-\frac{1}{1}log_2 \left(\frac{1}{1} \right) + \frac{-0}{1}log_2 \left(\frac{0}{1} \right) \right) = 0$$

Feature 1 (Summation):

$$E(H(1)) \sum_{i=1}^{10} \frac{p_i + n_i}{p + n} * \left(-\frac{p_i}{p_i + n_i}log_2 \left(\frac{p_i}{p_i + n_i} \right) + \frac{n_i}{p_i + n_i}log_2 \left(\frac{p_i}{p_i + n_i} \right) \right) = 0.4$$

(b) Now make the same computation, but if we created subsets using the second feature! (5pts)

$p_1 = 2 \qquad n_1 = 1$
$p_{-4} = 1 \qquad n_{-4} = 0$
$p_1 = 2 \qquad n_1 = 1$
$p_3 = 1 \qquad n_3 = 0$
$p_{11} = 1 \qquad n_{11} = 0$
$p_5 = 0 \qquad n_5 = 1$
$p_0 = 0 \qquad n_0 = 1$
$p_{-1} = 0 \qquad n_{-1} = 1$
$p_{-3} = 0 \qquad n_{-3} = 1$
$p_1 = 2 \qquad n_1 = 1$

$$E(H(A)) \sum_{i=1}^{K} \frac{p_i + n_i}{p + n} H \left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i} \right)$$

Feature 2 (first index):

$$E(H(1)) \sum_{i=1}^{1} \frac{3}{10} * \left(-\frac{2}{3}log_2 \left(\frac{2}{3} \right) + \frac{-1}{3}log_2 \left(\frac{1}{3} \right) \right) = 0.275$$

Feature 2 (second index):

$$E(H(1)) \sum_{i=2}^{2} \frac{1}{10} * \left(-\frac{1}{1}log_2 \left(\frac{1}{1} \right) + \frac{-0}{1}log_2 \left(\frac{0}{1} \right) \right) = 0$$

Feature 2 (Summation):

$$E(H(2)) \sum_{i=1}^{10} \frac{p_i + n_i}{p + n} * \left(-\frac{p_i}{p_i + n_i} log_2 \left(\frac{p_i}{p_i + n_i} \right) + \frac{n_i}{p_i + n_i} log_2 \left(\frac{p_i}{p_i + n_i} \right) \right) = 0.826$$

(c) Which feature is more discriminating based on results in Part (a) (5pt)?
   The first feature is more discriminating

(d) What are the principle components of the matrix X? For this (and the next) part you may assume that the features are *continuous*. Make sure the principle components are all unit length. You **MAY** use a utility function like *eig* or *svd* to determine these. (5pts)
$$cov_{x,y} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{N-1}$$
   To get the principle components, we need the eigen vectors, so we get them after using our original matrix to compute the covariance matrix. The resulting eigenvectors are:
$$\begin{bmatrix} -0.71636104 & 0.69772979 \\ -0.69772979 & -0.71636104 \end{bmatrix}$$

(e) If we were to project our data down to 1-D using the principle component, what would the new data matrix X be? (5pts)

   by multiplying the standardized data by our eigenvectors, the new data matrix would be
$$X = \begin{bmatrix} -0.1117808 \\ 0.23419233 \\ -0.27394203 \\ -0.12044108 \\ -2.74966176 \\ -0.77774625 \\ 0.54119423 \\ 1.35633049 \\ 0.71634587 \\ 1.18550899 \end{bmatrix}$$
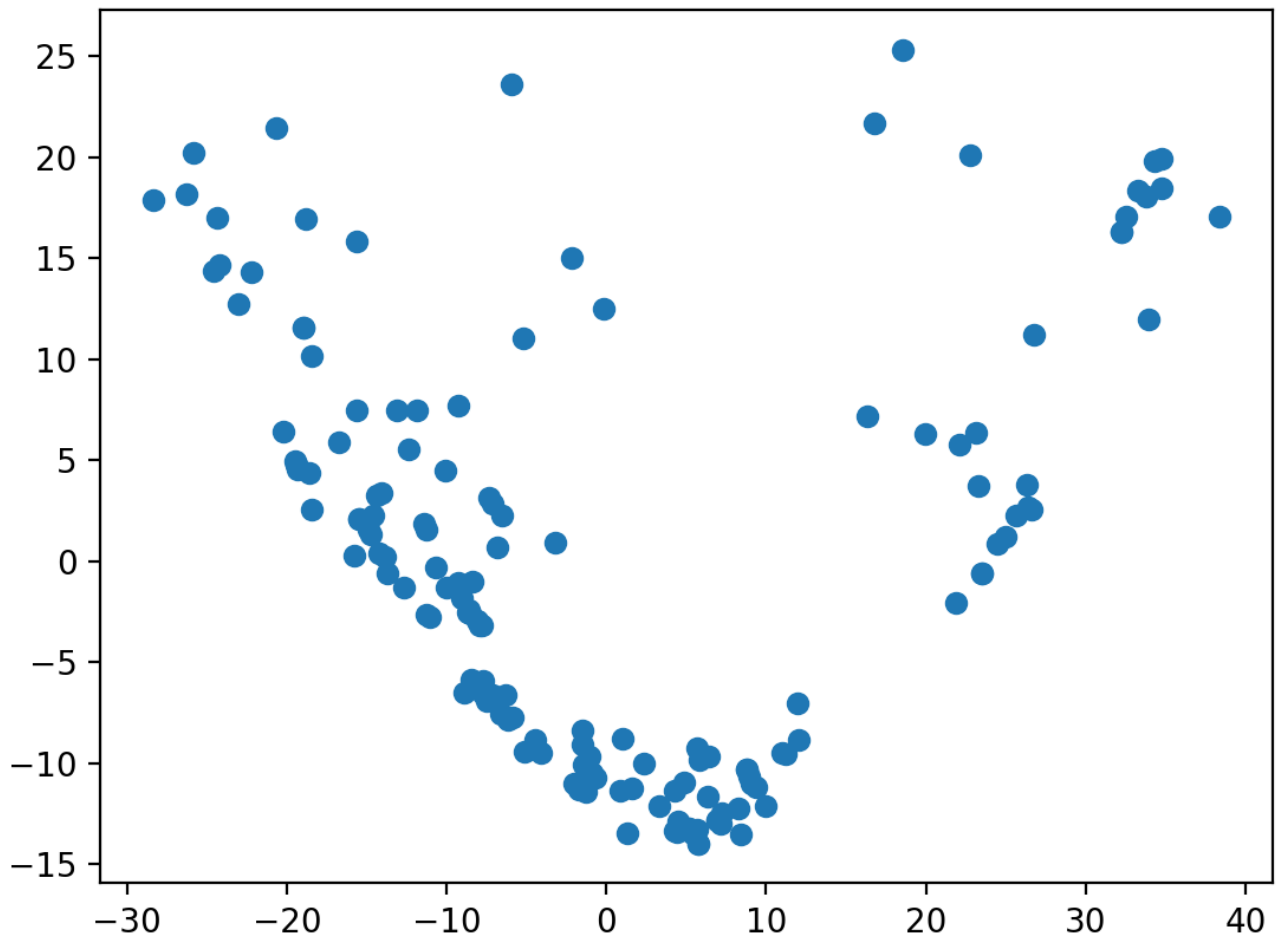
2 (30pts) Dimensionality Reduction via PCA

Figure 1: PCA Visualization