Your Name: CS74/174 Homework #1

Machine Learning and Statistical Data Analysis: Winter 2016

Problem 0: Getting Connected

Please join our class forum on Piazza,
`https://piazza.com/dartmouth/winter2016/cosc07401cosc17401wi16/home`
Please post your questions and discussion points on Piazza, rather than by email to me or the TA, since chances are that other students have the same or similar questions, and will be helped by seeing the discussion.

Problem 1: Matlab & Data Exploration

(a) Download and load the "Fisher iris" data set into Matlab (or Octave):

```
iris=load('data/iris.txt');     % load the text file
y = iris(:,end);           % target value is last column
X = iris(:,1:end-1);       % features are other columns
whos                       % show current variables in memory and sizes
```

The Iris data consist of four real-valued features used to predict which of three types of iris flower was measured (a three-class classification problem).

(b) Use `size(X,2)` to get the number of features, and `size(X,1)` to get the number of data points.

(c) The histogram ("`hist`") of the data is shown in Figure 1.

(d) Compute the mean of the data points for each feature (`mean`)

(e) The standard KNN is based on Euclidean distance $d(x^{(1)}, x^{(2)}) = \sqrt{\sum_i (x_i^{(1)} - x_i^{(2)})^2}$.

$$y = [y^{(1)}, \ldots, y^{(n)}] \tag{1}$$

$$X = [x^{(1)}, \cdots, x^{(n)}] \tag{2}$$

$$x^{(1)} = [x_1^{(1)}, \ldots, x_d^{(1)}]$$

$$X = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & x_3^{(1)} \\ x_1^{(2)} & x_2^{(2)} & x_3^{(2)} \\ x_1^{(3)} & x_2^{(3)} & x_3^{(3)} \\ x_1^{(4)} & x_2^{(4)} & x_3^{(4)} \end{bmatrix}$$

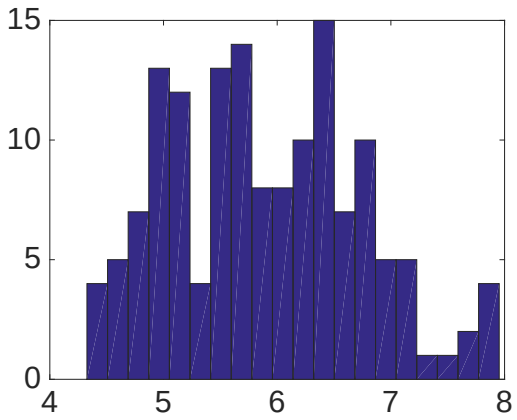

Figure 1: **Problem 1(c)**: The histogram of Iris data (the first feature)