CS273a Homework #1

Introduction to Machine Learning: Fall 2013

Due: Thursday October 3rd, 2013

Write neatly (or type) and show all your work!

This may be your first homework using MATLAB ; please see the course webpage for links to tutorials to help you start using it. In matlab, "help <functionname>" is often a good idea; e.g., "help plot", "help find"...

For MATLAB reports such as this one, I recommend that you use Word, OpenOffice, or LaTeX to create a document with your answers and any code snippets that will tell us what you did, export the finished report as a PDF file, and upload it to EEE. It is important that you include enough detail that we know how you solved the problem, since otherwise we will be unable to grade it.

In Windows or Mac, you can import MATLAB figures into your document directly using copy-/paste operations, or you can export a figure from MATLAB directly to pdf using e.g. "print -dpdf <filename>". You can also use this to create and import JPEG or PNG files, but if you do so, please ensure that they are of sufficient resolution to be clear in the resulting document ("-r<value>" controls the resolution).

I prefer to receive a *single* electronic document when possible; for some homeworks you may find it more convenient to do parts by hand, in which case a hard copy of that portion, or of the entire thing, is fine. Please *do not* upload a ZIP file containing multiple figures, as this is difficult to interpret without associated comments. If you need to submit large amounts of code, you can upload it in a ZIP file, but please ensure that there is enough information in your report to understand what you did without examining the code itself.

Problem 1: Getting Connected

Please join our class forum on Piazza,
`http://piazza.com/uci/fall2013/cs273a`
Please prefer to post questions and discussion points on Piazza, rather than by email to me, since chances are that other students have the same or similar questions, and will be helped by seeing the discussion.

Problem 2: Photos

Since there are a lot of us, please upload a recognizable photo of yourself to the EEE folder "Photos".

Problem 3: Matlab & Data Exploration

In this problem, we will explore some basic statistics and visualizations of an example data set. First, download and load the "Fisher iris" data set into Matlab:

```
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).

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

(b) For each feature, plot a histogram ("`hist`") of the data values

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

(d) Compute the variance and standard deviation of the data points for each feature

(e) "Normalize" the data by subtracting the mean value from each feature, and dividing by its standard deviation. (This will make the data zero-mean and unit variance.) Show your code. Note: you can do this with a for-loop (easy, but slow in Matlab), or in a "vectorized" form using `repmat` or `bsxfun` (faster, but harder to read).

(f) For each pair of features (1,2), (1,3), and (1,4), plot a scatterplot ("`plot`") of the feature values, colored according to their target value (class). (For example, plot all data points with $y = 0$ as blue, $y = 1$ as green, etc.) You may find the commands "`find`" and "`hold on`" useful for this. Alternatively, the command "`scatter`" may be useful.

Problem 4: Plotting functions

To plot a function in Matlab, you simply choose a selection of x locations, compute their associated y values, and plot it as a set of data points connected by lines.

(a) Plot feature 1 (x-axis) versus feature 3 (y-axis).

(b) To get the x and y range values, use `ax=axis()`

(c) Create a collection of densely-spaced x points, e.g., `xv = xmin:.01:xmax;`

(d) Evaluate your function at each of these points, e.g.,

```
yv = 1.5 * (xv - 3);
```

(e) Plot this line in blue on top of your data points. To make the line thicker, you may want to use options, `"linethickness",3` , in your `plot` call.

Problem 5: Writing a function

Write a function `dist` that calculates the *vector* of Euclidean distances between a single vector x and a collection of data points (stored in a matrix) X, e.g.,

$$D_j = \sqrt{\sum_i \left(x_i - X_{ji}\right)^2} \qquad \forall j$$

Your function should have interface:

```
function D = dist(x,X)
```

Use your function to calculate and show a histogram of the distances of the Iris data points from the first point, $X(1,:)$. Also include a listing of your function in your report.