CS 613 - Machine Learning

Assignment 1 - Dimensionality Reduction & Clustering

Alex Lapinski

Fall 2016

10/01/2016

Part 1 - Answers to Theory Questions

1. Why do we like to use quadratic error functions (say over a 4th degree polynomial function) (2pts)?

   The primary reason for using a quadradic error function is that is more easily tunable.

   $$y = ax^2 + bx + c \tag{1}$$

   In the equation above we are able to tune the a parameter, the b parameter and the c parameter to vary the placement and the shape of the quadratic plot. This allows us to easily define the location and shape of the plot.

2. Consider the following data:

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

(a) Find the principle components of the data (you must show the math, including how you compute the eivenvectors and eigenvalues). Make sure you standardize the data first and that your principle components are normalized to be unit length (5pts).

$Mean = \begin{bmatrix} -0.9 & 1.4 \end{bmatrix}$
$StandardDeviation = \begin{bmatrix} 4.228212 & 4.273952 \end{bmatrix}$

$$StandardizedData = (Data - Mean)/StandardDeviation = \begin{bmatrix} -0.260157 & -0.093590 \\ -0.969677 & -1.263468 \\ -0.496664 & -0.093590 \\ 0.212856 & 0.374361 \\ -1.679197 & 2.246165 \\ -0.260157 & 0.842312 \\ 0.449363 & -0.327566 \\ 1.395389 & -0.561541 \\ -0.023651 & -1.029492 \\ 1.631895 & -0.093590 \end{bmatrix}$$

(b) Project the data onto the principal component corresponding to the largest eigenvalue found in the previous part (3pts).

3. Consider the following data:

$$\text{Class 1} = \begin{bmatrix} -2 & 1 \\ -5 & -4 \\ -3 & 1 \\ 0 & 3 \\ -8 & 11 \end{bmatrix}, \text{Class 2} = \begin{bmatrix} -2 & 5 \\ 1 & 0 \\ 5 & -1 \\ -1 & -3 \\ 6 & 1 \end{bmatrix}$$

(a) Compute the information gain for each feature. You could standardize the data overall, although it won't make a difference. (5pts).

(b) Which feature is more discriminating based on results in part a (1pt)?

(c) Using LDA, find the direction of projection (you must show the math). Normalize this vector to be unit length.
Note: You don't not have to standardize the data since your computations should take into account the mean and standard deviations of the classes separately. (5pts).

(d) Project the data onto the principal component found in the previous part (3pts).

(e) Does the projection you performed in the previous part seem to provide good class separation? Why or why not (1pt)?

Part 2 - PCA Result

TODO: Include graph of visualization of the PCA Result

Part 3 - Visualization of k-means

Initial Setup

TODO: Insert Graph of initial setup visualization

Initial Cluster Assignment

TODO: Insert graph of initial cluster assignment visualization

Final Cluster Assignment

TODO: Insert graph of final cluster assignment visualization

Results

TODO: REport how many iterations it took for the algorithm to terminate

Raw Graphs

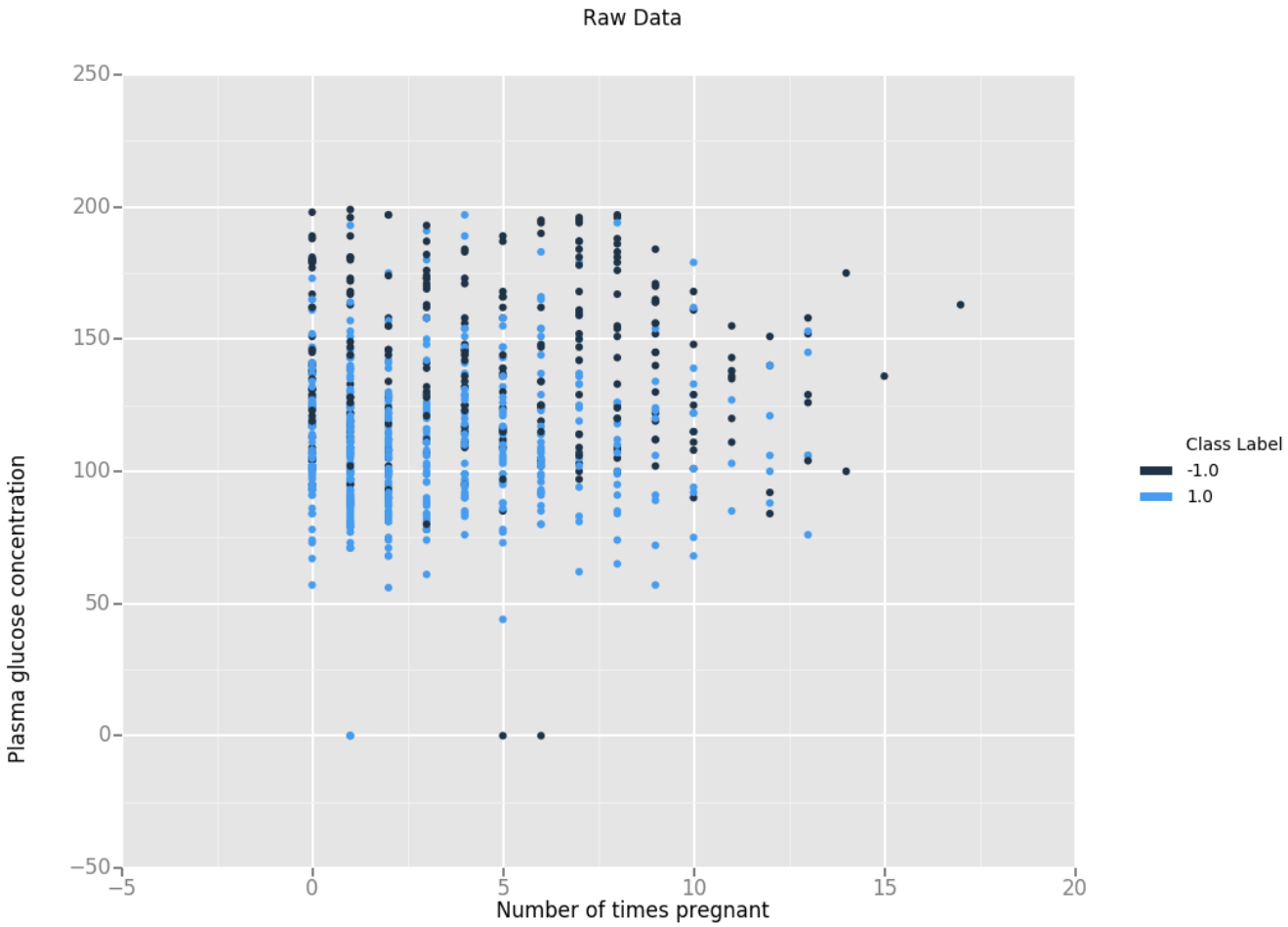

Figure 1: Raw Data 1

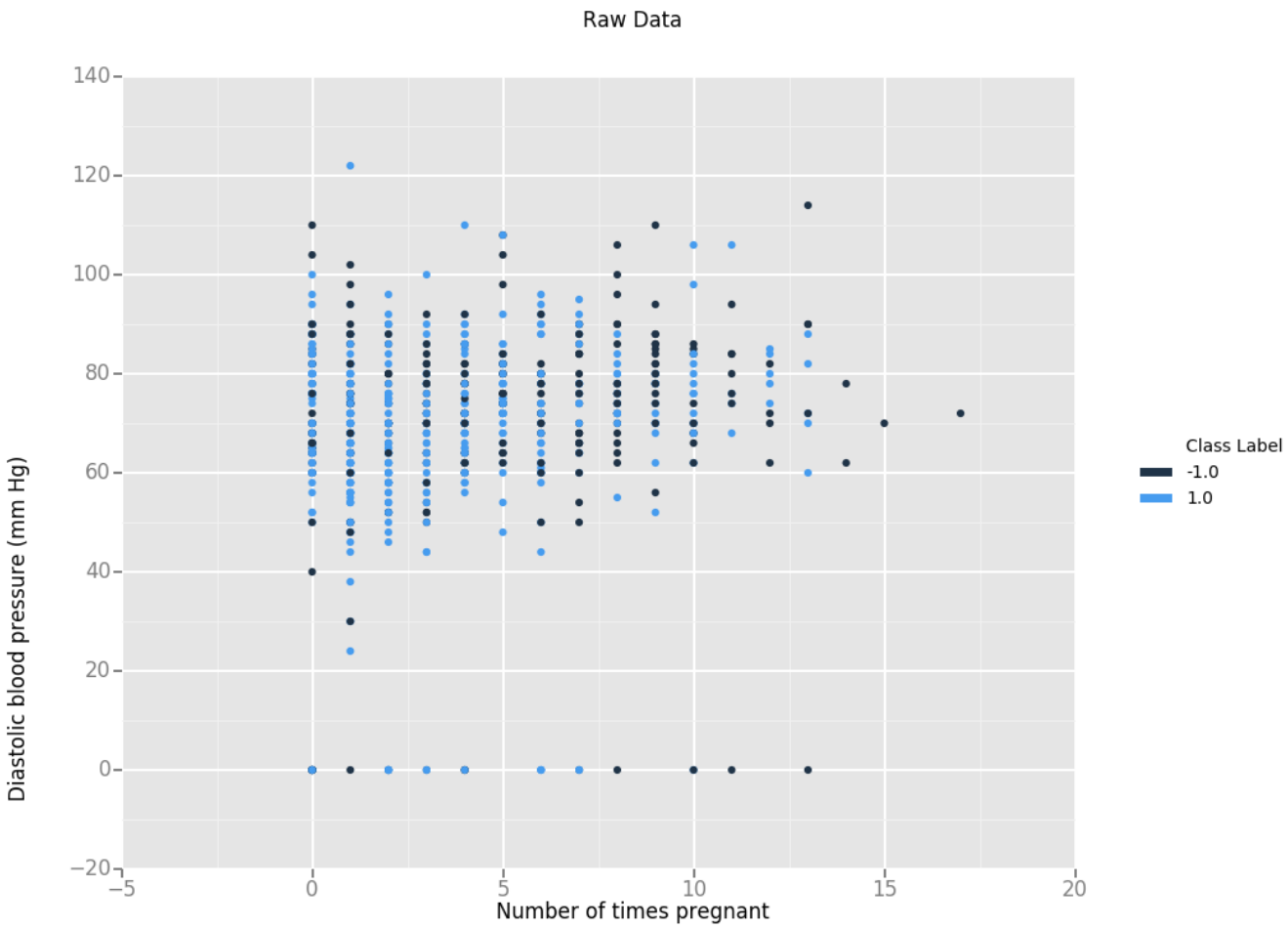

Figure 2: Raw Data 2

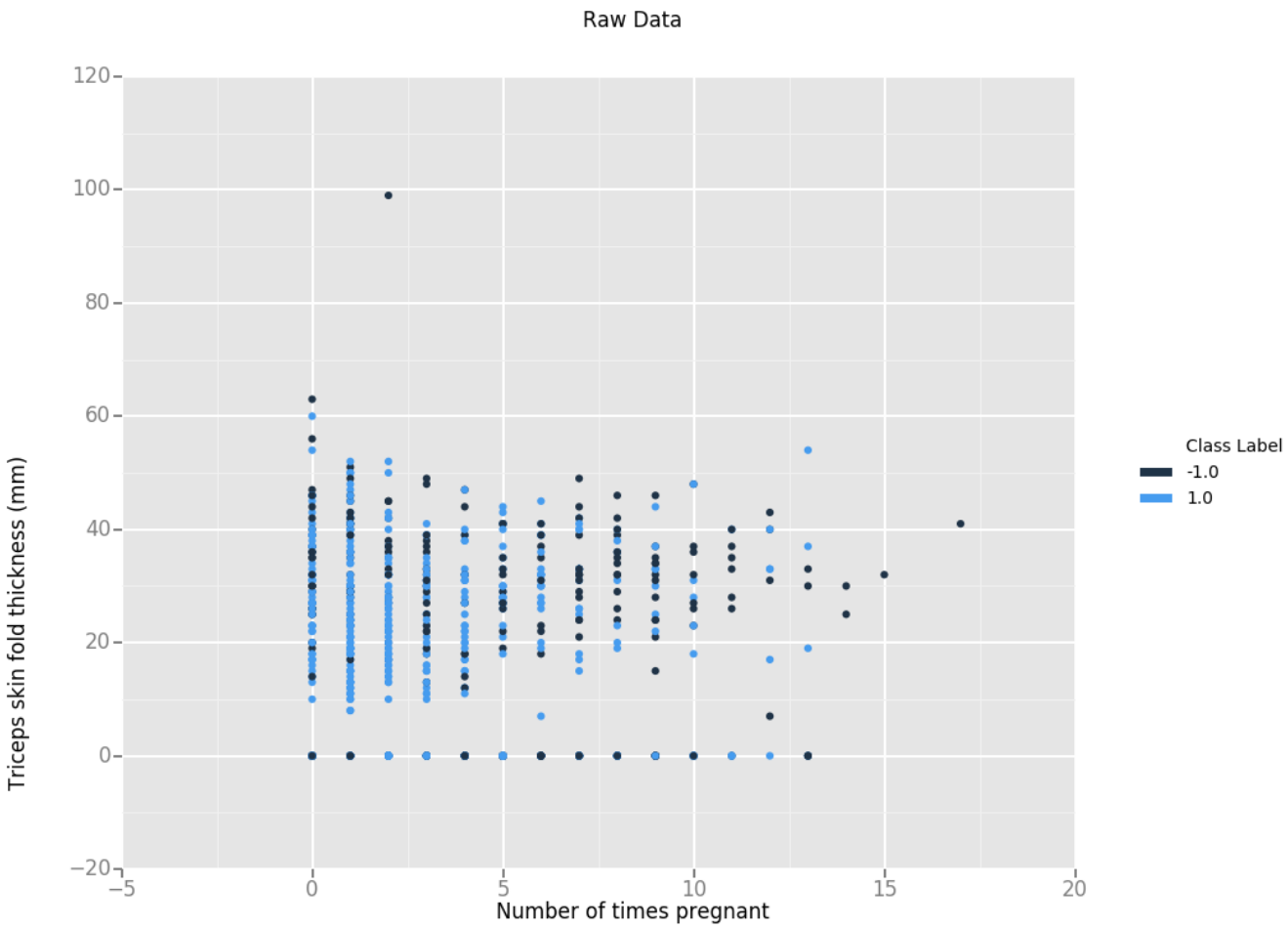

Figure 3: Raw Data 3