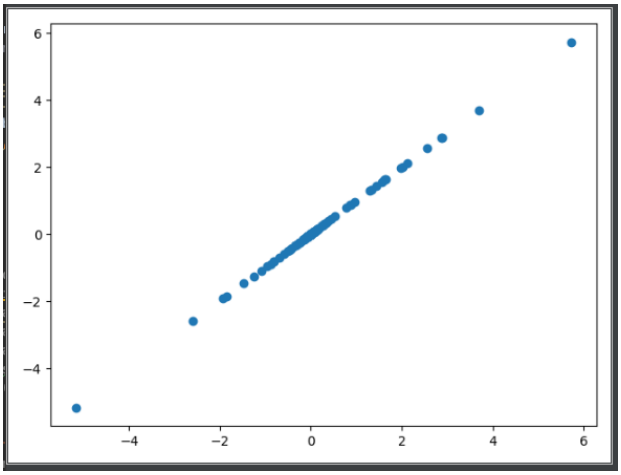
Writeup:

Shivansh Suhane
Ss4328
CS383: Machine Learning

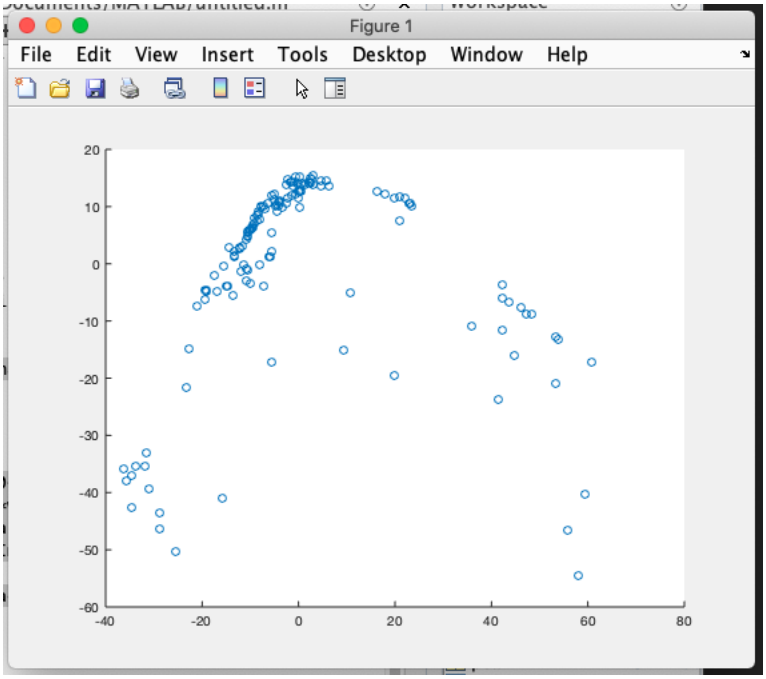
Q1:

a) In matlab, we form two classes to represent the data in each feature,
   class1 = [-2 1; -5 -5; ……] class2 = [-2 5; 1 0; 5, -1 ……..]
   Then we calculate the entropy using by using the formula from slides
   Entropy using class 1 is: 0.2000
b) Entropy becomes: 0.2755
c) First feature has greater entropy than second feature. More entropy means that the feature with more entropy will have more diverse dataset, and would therefore be more discriminating.
d) X= [2 1
   5 4
   3 1
   0 3
   8 11
   2 5
   1 0
   5 1
   1 3
   6 1];
   Principal components are [0.70 0,70
   0.70 -0.70]
e) The data becomes:
   [0.48,-1.00,1.18,0.51,11.83,3.35,-2.32,-5.85,-3.08,-5.10]

Q2) Tried this question in both python and Matlab. My python code produces a wrong plot, but my Matlab code produces the right plot.



Python scatter plot (wrong)



Matlab scatter plot

I think I'm choosing the wrong eigenvector for python but I've trouble validating this claim.

Q3)

I tried to continue in python, that didn't work.
Then I learned whatever matlab I could in one day and replicated everything in matlab and the q3 video generation still didn't work.
I put in more than 16 hours of effort and I still don't know how to proceed in 3.