1.c.

* Run “SVM\_implementation.m” (note time, ~0.000985 seconds)
* Values in command window – Optimal b: -16.31; Optimal w: 2.54, 3.07

1.d.

* Run “logRegression\_implementation.m” (note time, ~0.001993 seconds). Almost twice as slow as SVM implementation; SVM is better.
* If you run 1.d. right after 1.c. the decision boundaries will overlap. They have slightly different slopes and intercepts.

1.e.

* See Figure “Centroid and midline determination w/overlapped decision boundary”. When overlapping the plots, it does not appear to lie on the decision boundary in step e.

1.f.

* Given a known midpoint of the line that joins the centers of gravity, you can develop an algorithm that determines the normal to the midline at the center point. This normal is always orthogonal to the line, and can be easily calculated as the negative inverse slope of the midline. When you have the midpoint and you have the slope, you can determine values for the intercept using the equation of a line, as you have a slope and a point which lies on the line. Then you use this algorithm to make millions of dollars.

1.h.

* See Figure “Dataset including cluster and twofeature.txt dataset”.

1.i.

* Values in command window – Optimal b: -16.32; Optimal w: 2.54, 3.07. Numbers are nearly exact to Step 1.c. They are the same because the cluster does not impact the functional margin between the closest datapoints to the line, which are always maximized. The points considered are always the minimal of all geometric margins, therefore the cluster points are never considered as they would have the highest geometric margins.

2. A positive and negative support vector will always be produced because it is always possible to draw at least one hyperplane passing through one positive example, and another through one negative training example, i.e. two curbs of the road. SVM aims to maximize the distance between these two hyperplanes at all times, therefore a support vector can always be produced for at least one positive and one negative example.

Run “email\_classification.m” for remainder of Exercise 7.