Nguyen, Theodore

704-156-701

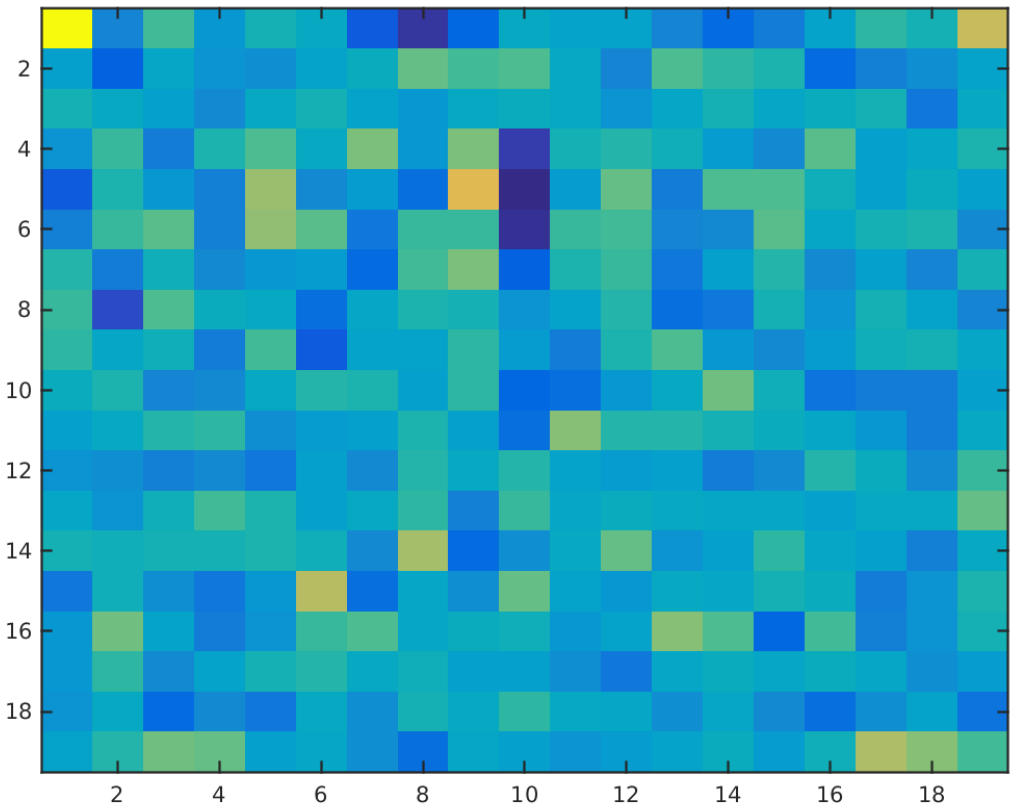
Math 156, spring 2016

Lab 7: Support Vector Machines

1. (5) Implement the soft support vector machine.

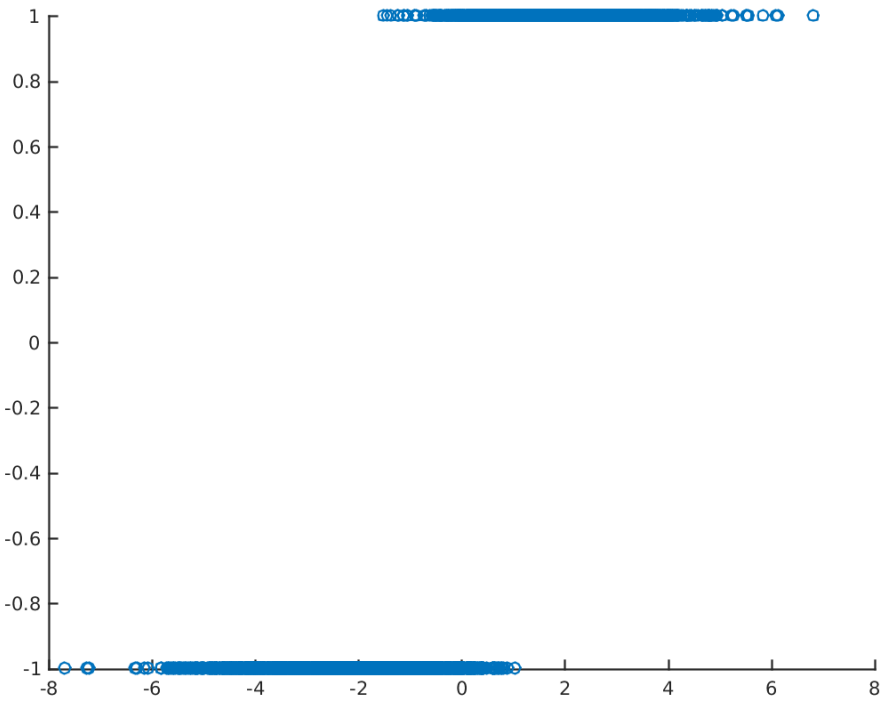
--- attached as code

2. (4) Load the CBCL dataset (check for dimensions of X, labels in L) and apply the soft SVM classi_er with a penalty = 0:005. Generate and turn in a visualization of w, as found by the SVM function, using the command imagesc(reshape(w, dims)) (here dims comes from the original data _le). Explain what you see.



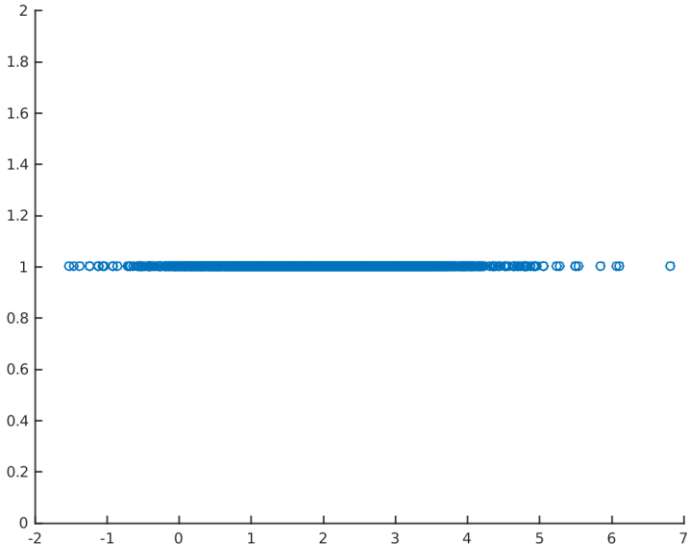
What we see is the visualization of the extremes. Blocks with extreme colors; that is, bright yellow or dark blue, are those with either a maximum or a minimum value for the values in w. Essentially, these values are given more weight

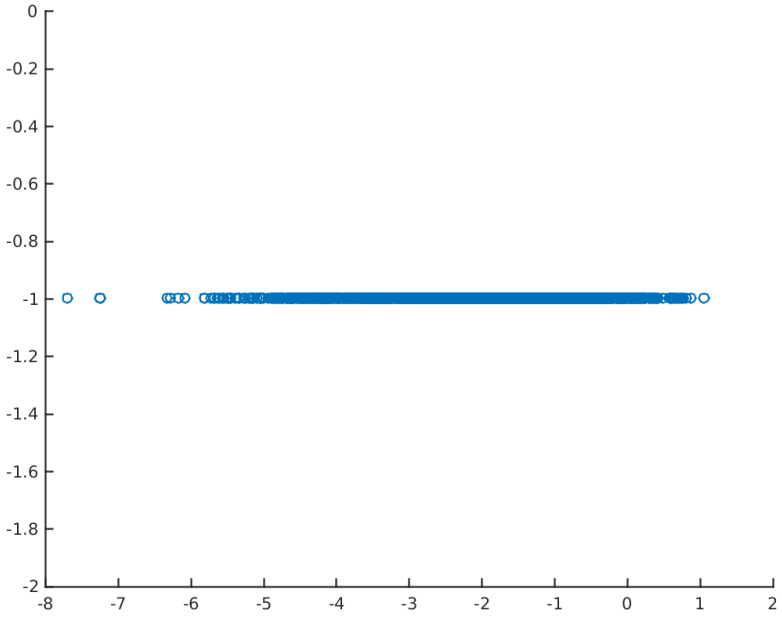
3. (4) Generate a plot of X'*w + b against L, the correct labels. What do the extremes (minimum/maximum) of this plot represent? Were any data points classi_ed incorrectly, and how can you tell?



Yes, data points were classified incorrectly; those that have a y-value of 1 and an x-value less than zero, as well as those with a y-value of -1 and an x-value greater than zero are misclassified. Essentially, this is a plot of SVM classification on the x-axis, and true classification on the y-axis. The signs must match up for the classification to be a success.

4. (4) Turn in two images corresponding to the extreme points of this plot, and two more images corresponding to example support vectors from each class.





5. Load the 20 Newsgroups data set (check for dimensions of X, labels in L) and apply the soft SVM with = 0:005.

6. (4) By examination of the vector w, which words are the most important for separating the two classes of documents? Which words are most distinctly space-related? What about cryptography-related? Give at least _ve important words for each case.

--- words closest to separator: **gets, Swarthmore, hopefully, exists, habitable**

--- words at maximum, related to space: **space, moon, pat, orbit, sci, dc**

--- words at minimum, related to crypto: **clipper, encryption, sat, key, chip, crypto**

7. (4) Is the 20 Newsgroups data linearly separable? How do you know?

-- no; the convex hulls of the two classes have an intersection