

## Assignment 4

You are given an SVM code that classifies examples in one of 6 options for datasets:

1. A standard linearly separable set
2. a non-linearly separable set similar to xor
3. Random set, arbitrary label
4. Random set, labels determined by a rule.
5. A set of two spirals
6. An imbalanced set.

Use the command "SVMtrial" and choose the different datasets. Consider using the 3D rotate option to look at the figures created.

- 1) (30) Go over the code in the SVMtrial function handed to you, debug it using the various input sets and observe the values of the variables calculated inside the function. Answer the following questions
  - a. What is the role of the variable Lambda? describe the effect of changing the value of Lambda. Hint – read the documentation of the quadprog function.
  - b. Describe the input and output (in particular, describe the input H, use the terms learned in class) of the quadprog function and explain in basic terms what it does (no need to explain how it does it).
  - c. Describe how the bias is estimated
  - d. Write pseudocode describing the given SVM algorithm
- 2) (50) Use the fourth option for input data, and follow the instructions in the following questions
  - a. Write a function that separates the dataset generated into training and validation sets using the K-fold cross validation method described in the [link](#).
  - b. Write a function to evaluate the model created by SVMtrial. Use the model, and the validation sets created by the function from 3a. use K=3, what is the average accuracy of the model?

- c. Test different sizes of training sets (2,4,8,16...) while keeping the size of validation sets constant (50). Plot the average accuracy of the model as a function of the size of the training set. Explain your results.
- d. Set the training set to 10 samples, change the value of Lambda from very small values to very large values. Plot the average accuracy of the model as a function of Lambda, explain your results.

3) (20) Write the equation solved by the quadprog function, show that it is equivalent to the SVM problem with slack variables we showed in class (online sources are allowed)

4)(bonus) choose a different kernel to apply, implement it in the code and test the result for the various datasets. Include the plots in the report and explain the effects of changing the kernel on the solutions found.

### Administration

In your report, only include figures that are required to aid in your explanations.

Exercises requiring submission are done in groups of three-four and should be emailed to the TA by the deadline. Submission should include **One file only**:

RAR / ZIP file containing all matlab code files required by the exercise and a DOC / DOCX / PDF file including full answers, graphs and reports required by the exercise.

Hand-written pages that are placed in order in the file and scanned in high quality are allowed.

**Code submitted must run and produce all figures presented in your report.**

Figures should have a title, labels for axes, and a short explanation of the information in the figure, and the conclusions derived from it

Email title should include #assignment, and the submitting students full name and ID

For example, Email subject: "Assignment 3, submitted by Eric Kandel 01234787485 and Mitch Hedberg 06487693241 ".

**Assignments handed in not following these instructions will have their grades reduced.**