STATS 415 Homework 9

Due Thursday Dec 3, 2019

1. For this question, please use graph paper; you can print it for free from many websites, for example, www.printfreegraphpaper.com. (15 points per question)

   (a) Draw the hyperplane defined by $2X_1 - 2X_2 - 1 = 0$. Indicate the set of points satisfying $2X_1 - 2X_2 - 1 > 0$ with a "+" sign, and the set of points satisfying $2X_1 - 2X_2 - 1 < 0$ with a "−" sign.

   (b) Suppose your hyperplane is the optimal separating hyperplane for an SVM classifier fitted to some data, with the margin $m = \sqrt{2}$. Draw the margin lines.

   (c) What class label (+ or −, as defined above) does this SVM predict for the following points: (1,4); (1,1); (2, -5); (2, -1); (4,2)?

   (d) Suppose these five points were part of the training data, and their true labels, given in the same order, are −, −, +, +, −. Calculate the corresponding slack values (ξ_i's) for each of the five points.

2. This question uses the same `crabs` data used in Homework 8, with the five body measurements as predictors and the same response (species). Omit all other variables including sex, and use the same split into training and test data as in HW8. (20 points per question)

   (a) Fit a linear SVM to the data for a range of values of `cost`, to predict Species from the five numerical measurements. Plot the training, cross-validation, and test errors as a function of `cost`. Comment on your results and make some relevant plots.

   (b) Now fit nonlinear SVMs with radial and polynomial kernels, for a range of values of `gamma` and `degree` and `cost`. Plot the training, cross-validation, and test errors as a function of `cost`. Comment on your results.

.

Please limit your answer to Q2 to no more than 5 pages, organized into a coherent typed data analysis report. Answers to Q1 may be either typed or handwritten. Please submit your homework through Canvas.