Tutorial 11 Support vector machine

Lecture

Last week you are introduced the support vector machine

Textbook Reading

Chapter 9.1 to 9.5 of James et al.

Tutorial

The lectures last week introduced a new supervised method, namely support vector machine. The algorithm can be applied to both classification and regression problems, however we only covered the classification problem in this subject.

Labs

Exercise

In a group of two, discuss

1. What is a hyperplane?
2. What is the mathematical formulation of a hyperplane?
3. When does a hyperplane pass through the origin?
4. How to classify using a hyperplane?
5. What is separating hyperplane?
6. What is the maximal margin classifier?
7. How to construct the maximal margin classifier?
8. What is support vector?
9. What is the soft margin classifier?
10. How to construct the support vector classifier?
11. What is support vector machine?
12. SVM for K > 2
13. Similarity between SVM and logistic regression?

Chapter 9, Exercise 2

We have seen that in $p = 2$ dimensions, a linear decision boundary takes the form $\beta_0 + \beta_1 x_1 + \beta_2 x_2 = 0$.We now investigate a non-linear decision boundary

1. Sketch the curve
$$(1+x_1)^2 + (2-x_2)^2 = 4$$

2. On your sketch, indicate the set of points for which
$$(1+x_1)^2 + (2-x_2)^2 > 4$$
as well as the set of point for which
$$(1+x_1)^2 + (2-x_2)^2 \leq 4$$

3. Suppose that a classifier assigns an observation to the blue class if
$$(1+x_1)^2 + (2-x_2)^2 > 4$$
and to the red class otherwise. To what class is the observation (0,0), (-1,1), (2,2) and (3,8) classified?

4. Argue that while the decision boundary in (3) is not linear in terms of x_1 and x_2, it is linear in terms of x_1, x_1^2, x_2 and x_2^2

Chapter 9, Exercise 7

In this problem, you will use support vector approaches in order to predict whether a given car gets high or low gas mileage based on the `Auto` data set

1. Create a binary variable that takes on a 1 for cars with gas mileage above the median, and a 0 for cars with gas mileage below the median.

2. Fit a support vector classifier to the data with various values of cost, in order to predict whether a car gets high or low gas mileage. Report the cross-validation errors associated with different values of this parameter. Comment on your results

3. Now repeat (2), this time using SVMs with radial and polynomial basis kernels, with different values of cost and degree and cost. Comment on your results.

4. Make some plots to back up your assertions in (2) and (3).

   Hint: In the lab, we used the `plot()` function for `svm` objects only in cases with $p = 2$. When $p > 2$, you can use the `plot()` function to create plots displaying pairs of variables at a time. Essentially, instead of typing

   ```
   plot(svmfit , dat)
   ```

   where svmfit contains your fitted model and dat is a data frame containing your data, you can type

   ```
   plot(svmfit , dat , x1∼x4)
   ```

   in order to plot just the first and fourth variables. However, you must replace x_1 and x_4 with the correct variable names. To find out more, type `?plot.svm`