Solution for Homework 2

Oct 7, 2018

Problem 1

Part 1

Similar to the first assignment, I use quadprog to solve the primal SVM. In order to solve the primal SVM, I change the parameters for quadprog to fit out problem. I attach my code to show the mapping.

Accuracy for different values of C for the training set:

$C = 1$	$C = 10$	$C = 100$	$C = 1000$	$C = 10^4$	$C = 10^5$	$C = 10^6$	$C = 10^7$	$C = 10^8$
88.462	88.462	88.462	89.744	88.462	89.744	94.872	97.436	98.718

Accuracy for different values of C for the validation set:

$C = 1$	$C = 10$	$C = 100$	$C = 1000$	$C = 10^4$	$C = 10^5$	$C = 10^6$	$C = 10^7$	$C = 10^8$
86.207	86.207	86.207	81.034	81.034	84.483	81.034	82.759	81.034

For C =1, 10 and 100, in validation set, we get the accuracy equals to 86.207 for each of them. So the accuracy on test set for these C is as follows:

Accuracy for selected classifier on the test set:

$C = 1$	$C = 10$	$C = 100$
83.051	81.356	84.746

Part 2

Similar to the first assignment, I use quadprog to solve the dual SVM. In order to solve the dual SVM, I change the parameters for quadprog to fit out problem. I attach my code to show the mapping.

Accuracy for different values of c and sigma for the training set:

σ \ C	$C = 1$	$C = 10$	$C = 100$	10^3	$C = 10^4$	$C = 10^5$	$C = 10^6$	$C = 10^7$	$C = 10^8$
$\sigma = 0.1$	100	100	100	100	100	100	100	100	100
$\sigma = 1$	100	100	100	100	100	100	100	100	100
$\sigma = 10$	93.59	97.436	100	100	100	100	100	100	100
$\sigma = 100$	87.179	87.179	89.744	97.436	75.64	65.385	65.385	65.385	65.385
$\sigma = 1000$	78.20	78.20	78.20	83.333	87.179	66.667	58.974	58.974	58.974

Accuracy for different values of c and sigma for the validation set:

σ \ C	C = 1	C = 10	C = 100	C = 10^3	C = 10^4	C = 10^5	C = 10^6	C = 10^7	C = 10^8
$\sigma = 0.1$	74.138	74.138	74.138	74.138	74.138	74.138	74.138	74.138	74.138
$\sigma = 1$	74.138	74.138	74.138	74.138	74.138	74.138	74.138	74.138	74.138
$\sigma = 10$	82.759	81.034	79.310	79.310	79.310	79.310	79.310	79.310	79.310
$\sigma = 100$	79.310	77.586	81.034	86.207	75.862	63.793	63.793	63.793	63.793
$\sigma = 1000$	74.138	74.138	74.138	82.759	82.759	58.621	48.276	48.276	48.276

For c = 1000 and σ = 100 in validation set, we get the accuracy equals to 86.207 . So the accuracy on test set for these values is 79.661

Part 3

(a) The accuracy of the learned classifier on the training set is:

$$k = 1, \text{accuracy} = 1.0$$
$$k = 5, \text{accuracy} = 0.8333333333333334$$
$$k = 11, \text{accuracy} = 0.8205128205128205$$
$$k = 15, \text{accuracy} = 0.7692307692307693$$
$$k = 21, \text{accuracy} = 0.782051282051282$$

(b) The accuracy on the validation set is:

$$k = 1, \text{accuracy} = 0.7068965517241379$$
$$k = 5, \text{accuracy} = 0.8103448275862069$$
$$k = 11, \text{accuracy} = 0.8103448275862069$$
$$k = 15, \text{accuracy} = 0.7758620689655172$$
$$k = 21, \text{accuracy} = 0.7758620689655172$$

The max accuracy on validation set is: 0.8103448275862069.

(c) With the best k selected from part (b), the accuracy on the test set is:

$$k = 5, \text{accuracy} = 0.6779661016949152$$
$$k = 11, \text{accuracy} = 0.711864406779661$$

4. Comparison

Based on the results listed from the previous parts, the dual SVMs with Gaussian Kernels should be preferred for this classification task. Since all of them were using the same training set and validation set, but the SVMs with Gaussian Kernels has a better accuracy on the test set.

2 Poisonous Mushrooms

1 See next page.

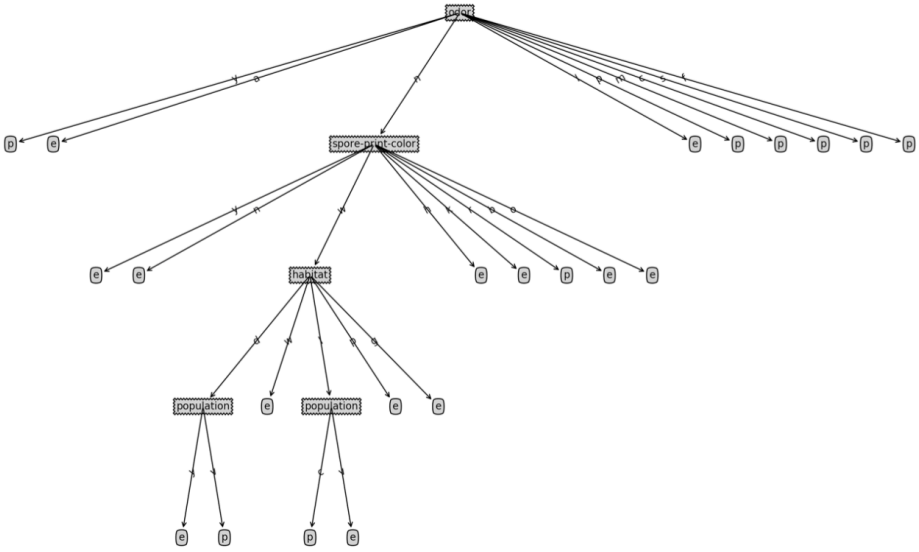
2 Size of the decision tree = 27 nodes.

3 Height of the decision tree = 4.

4 Accuracy on training set=100%.

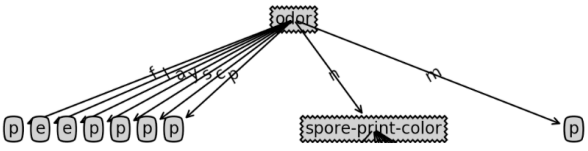
5 Accuracy on test set=100%.

6 For this problem, decision tree learning works really well, but it may not work just as well on other mushroom data. In that case, you can retrain the decision tree by combining that set of the data with current data and it will be guaranteed to perform well, as decision tree can represent any boolean function.



7 Since decision tree is able to learn any boolean function, it can overfit the training set. Thus, it is very much dependent on training/test split.

8 One non-leaf node decision tree:



Pick the root of learned decision tree above as the single non-leaf node, and use majority vote on the value "n" branch, after the calculation, the result will be the same. So, both of them are equal.