Solutions for Assignment 2

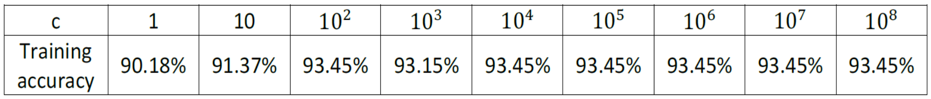
10-10-2017

1 Breast Cancer Diagnosis

1.1 Primal SVMs

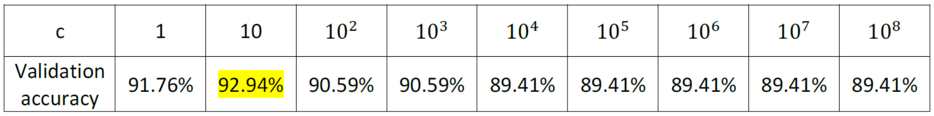
a To solve the SVM primal problem with slack formulation, you can use quadprog as stated in assignment 1 solutions. This time, the $\bar{w}$ will be $[w,\ b,\ \xi]$. quadprog's H arguement will be a diagonal matrix with coefficients 1s for w and 0 for b,ξ. You can similarly transform $y_i(w^T x^i + b) \geq 1 - \xi_i, \forall i$ into $A\bar{w} \leq b$ form for quadprog.

b Accuracy values for different values of c on training set are below:



c	1	10	10^2	10^3	10^4	10^5	10^6	10^7	10^8
Training accuracy	90.18%	91.37%	93.45%	93.15%	93.45%	93.45%	93.45%	93.45%	93.45%

c Accuracy values for different values of c on validation set are below:



c	1	10	10^2	10^3	10^4	10^5	10^6	10^7	10^8
Validation accuracy	91.76%	92.94%	90.59%	90.59%	89.41%	89.41%	89.41%	89.41%	89.41%

d When c = 10, the accuracy on test set is 96.62%.

1.2 Dual SVMs with Gaussian Kernels

a To solve Dual SVM with gaussian kernel, we can again use quadprog with proper parameters.
Dual SVM minimization problem is:

$$\min_{\lambda \geq 0} \frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j y_i y_j k(x^i, x^j) - \sum_i \lambda_i$$

$$\text{such that } \sum_i \lambda_i y_i = 0$$

$$0 \leq \lambda_i \leq c, \forall i$$

Code for solving this problem using quadprog is attached. By solving it, we get set of λ_i values. We know that $w = \sum_i \lambda_i y_i \phi(x_i)$. Since we don't know values of $\phi(x_i)$ but only kernel values, we can't get value of w. By complementary slackness, $0 < \lambda_k^* < c$ means $\phi(x^k)$ is support vector. Since $y_k(w^T\phi(x^k)+b) = 1$,

$$b = y_k^{-1} - w^T\phi(x^k)$$
$$b = y_k^{-1} - w^T\phi(x^k)$$
$$b = y_k^{-1} - \sum_i \lambda_i y_i \phi^T(x^i)\phi(x^k)$$
$$b = y_k^{-1} - \sum_i \lambda_i y_i k(x^i, x^k)$$

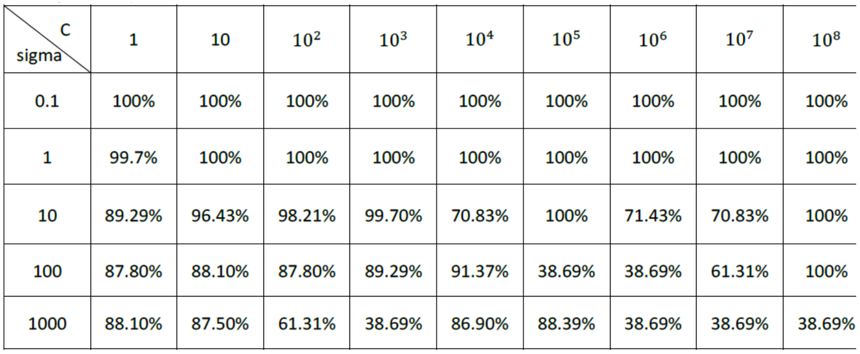
where, $0 < \lambda_k < c$

Using the equation derived, we can get value of b.
Now to classify given test data point x,

$$\begin{aligned} f(x) &= w^T\phi(x) + b \\ &= \sum_i \lambda_i y_i \phi(x_i)\phi(x) + b \\ &= \sum_i \lambda_i y_i k(x_i, x) + b \end{aligned}$$

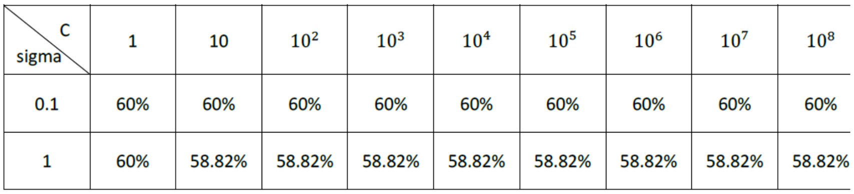
Thus, we can classify test set, even though we don't know values of w.

b Accuracy of trained classifier on training set for each value of c and σ is below:

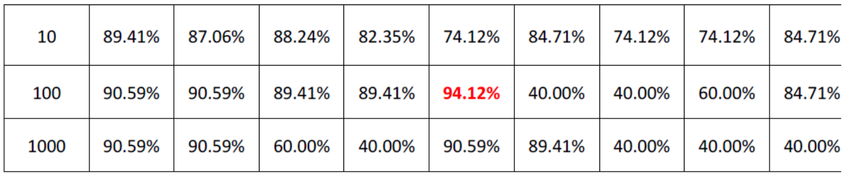


C / sigma	1	10	10²	10³	10⁴	10⁵	10⁶	10⁷	10⁸
0.1	100%	100%	100%	100%	100%	100%	100%	100%	100%
1	99.7%	100%	100%	100%	100%	100%	100%	100%	100%
10	89.29%	96.43%	98.21%	99.70%	70.83%	100%	71.43%	70.83%	100%
100	87.80%	88.10%	87.80%	89.29%	91.37%	38.69%	38.69%	61.31%	100%
1000	88.10%	87.50%	61.31%	38.69%	86.90%	88.39%	38.69%	38.69%	38.69%

c Accuracy of trained classifier on validation set for each value of c and σ is below:



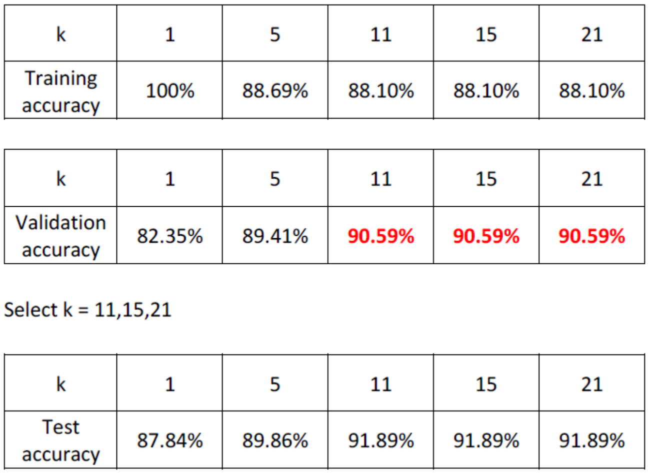
C / sigma	1	10	10²	10³	10⁴	10⁵	10⁶	10⁷	10⁸
0.1	60%	60%	60%	60%	60%	60%	60%	60%	60%
1	60%	58.82%	58.82%	58.82%	58.82%	58.82%	58.82%	58.82%	58.82%



10	89.41%	87.06%	88.24%	82.35%	74.12%	84.71%	74.12%	74.12%	84.71%
100	90.59%	90.59%	89.41%	89.41%	**94.12%**	40.00%	40.00%	60.00%	84.71%
1000	90.59%	90.59%	60.00%	40.00%	90.59%	89.41%	40.00%	40.00%	40.00%

d Select c = 10000, sigma = 100, C and sigma is different if you choose different support vector to calculate b. Test accuracy for this classifier is 93.92%

1.3 KNN



k	1	5	11	15	21
Training accuracy	100%	88.69%	88.10%	88.10%	88.10%

k	1	5	11	15	21
Validation accuracy	82.35%	89.41%	**90.59%**	**90.59%**	**90.59%**

Select k = 11,15,21

k	1	5	11	15	21
Test accuracy	87.84%	89.86%	91.89%	91.89%	91.89%

1.4 Preferred Classifier

From the accuracy computed using each approaches, I will choose prim SVM. Because the test accuracy for this model is the highest, and run more quickly.

2 Poisonous Mushrooms

1 See next page.

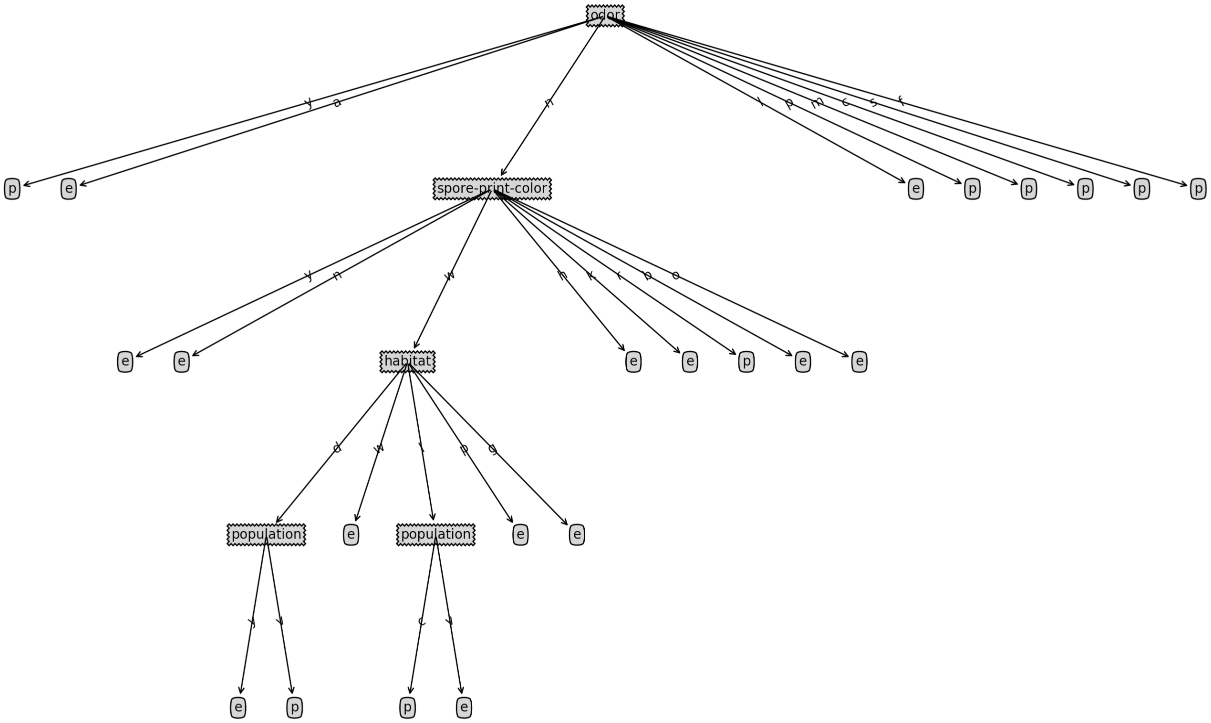
2 Size of the decision tree = 27 nodes.

3 Height of the decision tree = 4.

4 Accuracy on training set=100%.

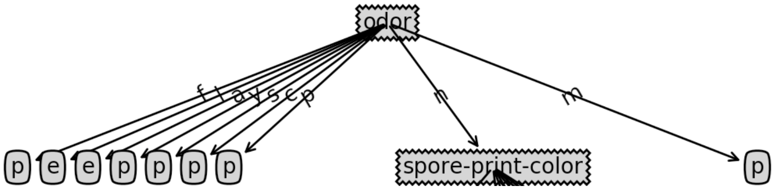
5 Accuracy on test set=100%.

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.