Cpts 570 Homework 2

Xinyu Liu

October 2021

Analytical Analysis

1

a

We know that $w =$ vector point from - to +. Therefore $w = C_+ - C_-$. Any point on the hyperplane $w \cdot x + b$ is 0 and hyperplane is in the middle of C_+, C_-. Therefore, we have $w \cdot \frac{C_+ - C_-}{2} + b = 0$. Use w we found, $b = -(C_+ - C_-) \cdot \frac{C_+ - C_-}{2}$

b

Number of support vectors = points that lies on margin

2

a

$exp(-\frac{1}{2}||x_i - x_j||^2)$

$exp(-\frac{1}{2}(||x_i||^2 - 2x_ix_j + ||x_j||^2))$

$exp(x_ix_j - \frac{1}{2}||x_i||^2 - \frac{1}{2}||x_j||^2)$

$exp(x_ix_j)exp(-\frac{1}{2}||x_i||^2 - \frac{1}{2}||x_j||^2)$

$exp(-\frac{1}{2}||x_i||^2 - \frac{1}{2}||x_j||^2) =$ some constant C

$C * exp(x_ix_j)$

$= C * \sum_{n=0}^{\infty} \frac{x_i^n x_j^n}{n!}$

$= C * \sum_{n=0}^{\infty} \frac{(x_ix_j)^n}{n!}$

$= C * \sum_{n=0}^{\infty} \frac{K(x_ix_j)}{n!}$

Kernel for any n

b

$||\phi(x_i) - \phi(x_j)||^2$

$= ||\phi(x_i)||^2 - 2\phi(x_i)\phi(x_j) + ||\phi(x_j)||^2$

RBF kernel has range (0, 1)

$= 1 - 2 * 0 + 1 = 2 \leq 2$

3

Again, RBF kernel has range (0,1), when x_i and x_j are very far, $K(x_i, x_j) = 0$

Therefore $f(x_{far}) = \sum a_i y_i K(x_{far}, x_i) + b = 0 + b$

4

According to Mercer's theorem, the kernel function has to be symmetric and positive-finite. We know that a legal kernel function is K $= < x_i, x_j >$, which means $< x_i, x_j >$ is positive. Therefore, $- < x_i, x_j >$ is not positive and does not satisfy Mercer's theorem.

5

Soft margin optimization is $min \frac{1}{2}||w||^2 + C \sum \xi_i$.

$C \sum \xi_i$ means bend for fitting training samples by C * sum of all cost, therefore we substitute ξ_i with C_+ and C_- depending on whether i is +1 or -1.

Changed version would be $min \frac{1}{2}||w||^2 + C \sum_i C_i$ where $C_i = C_+$ or C_-

6

a

Training formula would be optimizing $min \frac{1}{2}||w||^2 + c \sum_i \xi_i$ where ξ_i is the coarseness of i-th cluster

b

If current refine level accuracy is too low, then increase the amount of refinement.

c

The refining SVM should stop when current cluster accuracy = next level refined cluster accuracy

7

a

He only needs to pick B-th closest to margin support vectors from SV, because those have more weight values for classifying.

b

$f(u) = \sum_{i \in SV}^{B} \alpha_i x_i u$. for all B are obtained from (a), then update weights based on $f(u)$

Programming and Empirical Analysis

1

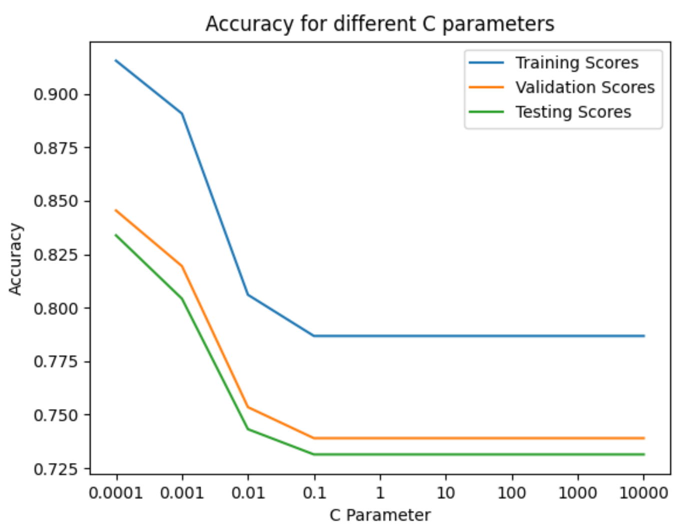
1a.

The graphs for 1a takes too long to generate since SVC() takes long to converge. However, I have accuracy for what I've got so far.

	Training	Validation	Testing
10^{-4}	0.91620...	0.8454...	0.8351
10^{-3}	0.9228...	0.83775	0.8282
10^{-2}	0.9267...	0.8355	0.827

We can see that as C parameter increases, so does training accuracy. However, validation and testing accuracy go down. When C is too large, that means tolerance for training data is too high, thus causing overfitting.

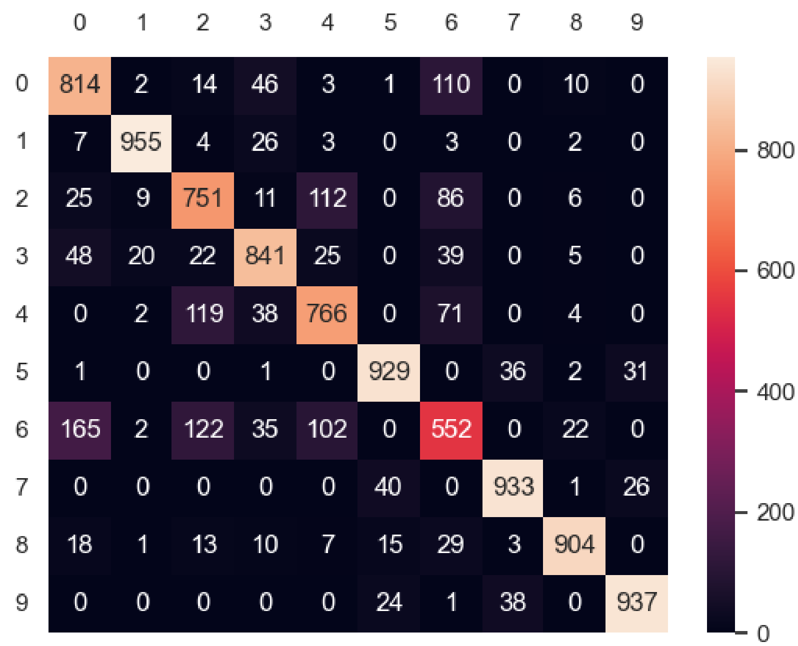
Following is the graph when max_iter = 500000. Since it doesn't reach convergence, we can't really say much about this result. However, we can see that accuracy becomes a flat line starting at 10^{-1}. It's possible that C becomes too large and completely overrules optimization.



1b.

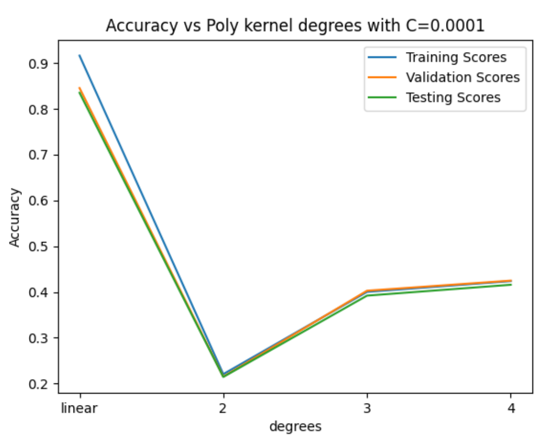
- Best C is: 0.0001
- Accuracy: 0.8382

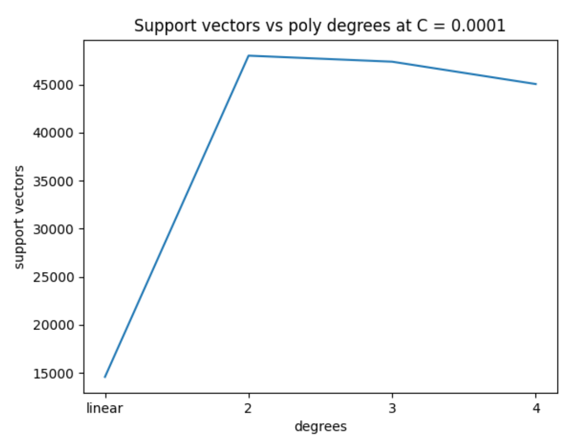
Confusion matrix:



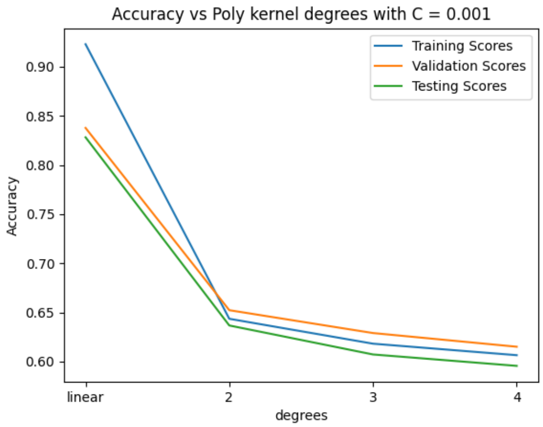
1c.

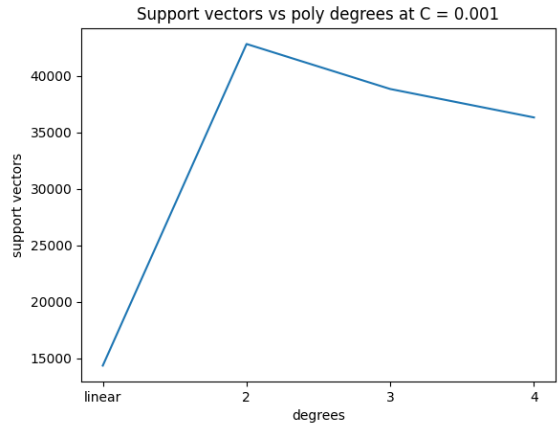
We find C = 0.0001 is the best option from 1a. When we use a polynomial kernel function of 2, 3, 4 degrees, we can see that accuracy drop dramatically. Therefore, we can conclude that linear kernel fits best for fashion-mnist data. Also, support vectors increase as accuracy drop, likely it's because when using extra support vectors for margin maximization causes overfitting.

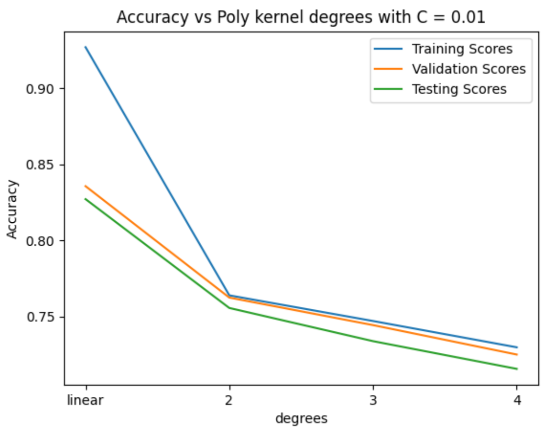


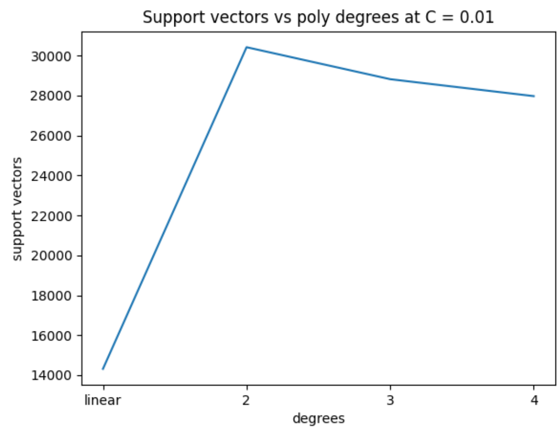


Following are some extra graphs for other C.









2

Graph takes way too long, therefore I only have code for this question.

3

	Training	Validation	Testing
ID3	1.0	0.9705	0.9489
Pruned	0.9916	1.0	0.9708

As we can see, the training accuracy for ID3 is 100% as expected. Validation and testing are relatively high also. After pruning, there is a big increase in accuracy for validation and testing with only a very slight drop for training.