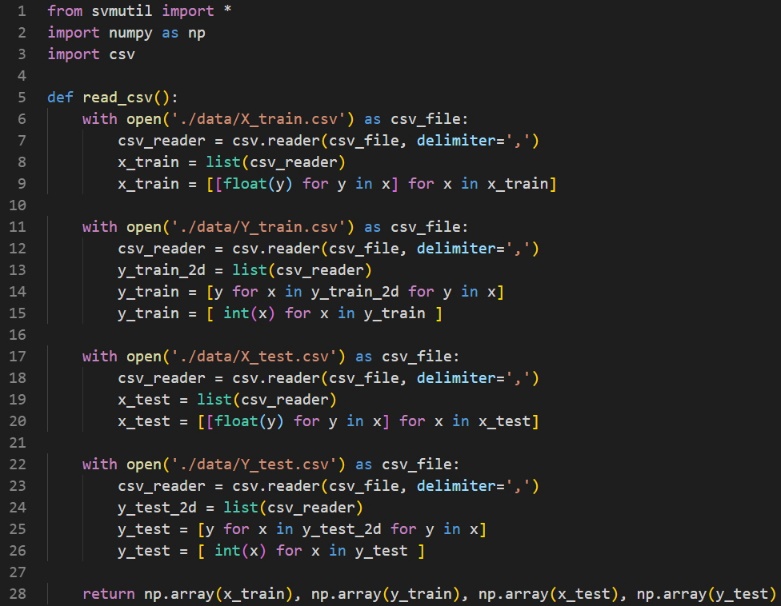
**Machine Learning HW5 Report by 0516075**

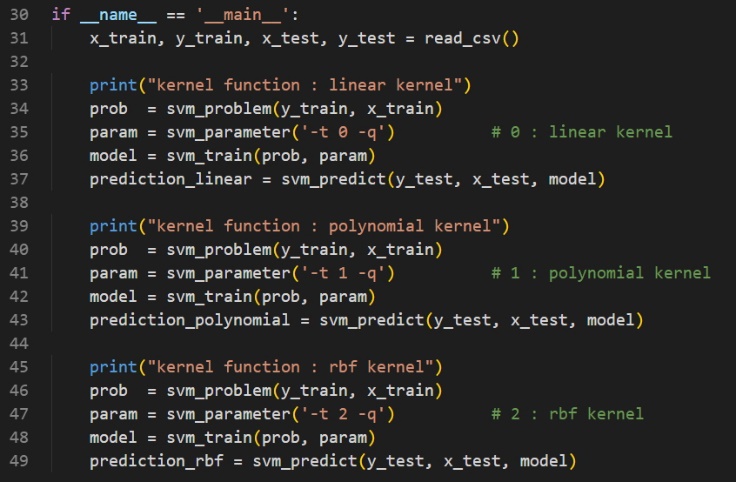
1. Gaussian Process
2. SVM on MNIST dataset



Function read\_csv is used to read all csv file and store them into numpy array form.Below parts all have this function to read X\_train, Y\_train, X\_test, and Y\_test.

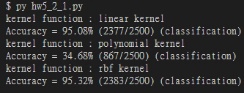
2-1.

The code of this part is related to hw5\_2\_1.py



In this part, we first use function read\_csv to read csv files and store as numpy arrays.Lines 34th, 40th, 46th use svm\_problem in libsvm to define our problem sets.

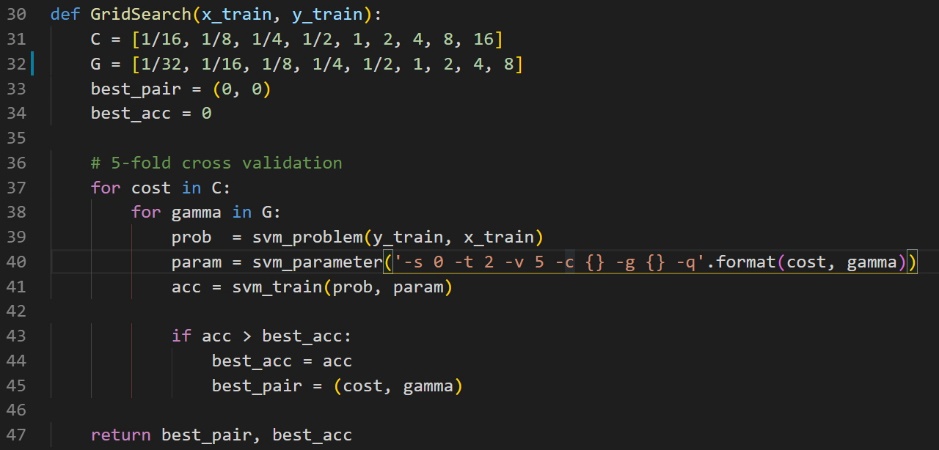
Lines 35th, 41th, 47th can custormize parameter we want by using svm\_parameter.”-q” means silent mode.”-t” means kernel\_type.Parameter 0 of t means linear kernel, 1 means polynomial kernel, 2 means RBF kernel.



As we can see, polynomial kernel only has 34.68% performance, while linear kernel and RBF kernel get about 95% accuracy.RBF kernel’s performance is 95.32%, which is a little higher than linear kernel.Note that other parameters of these kernels use default value.

2-2.

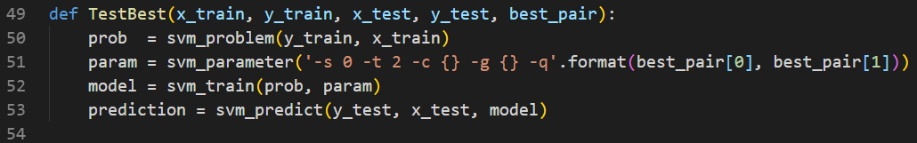
The code of this part is related to hw5\_2\_2.py.



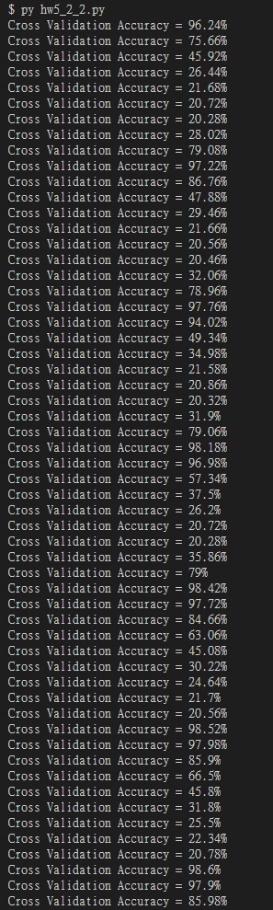
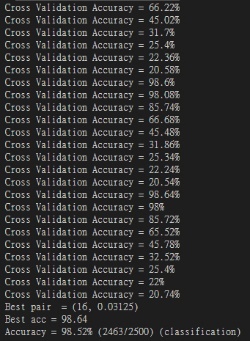
When finishing read\_csv function, we then use training data as GridSearch function’s input.I select C and γ by reference [1].It suggests using exponentially growing sequences of C and γ is a pratical method to identify good parameters.Because of time cost, I only select 9 values of each to do grid search.In line 40th, we also use quiet mode “-q”, -s means svm\_type, parameter 0 means C-SVC.”-t” is same as before, means kernel\_type. In 2-1 we have discovered that rbf kernel have best performance of three kernel types, so here I set parameter of kernel\_type is 2.

”-c” means to modify cost parameter in C-SVC, “-g” means to modify gamma parameter.

“-v” means cross validation, here I want to use 5-fold cross validation so the parameter of –v is 5.

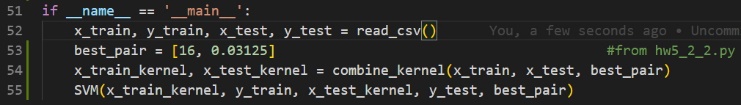


After tuning best parameter, function TestBest is to select best parameter as training parameter to predict testing data.

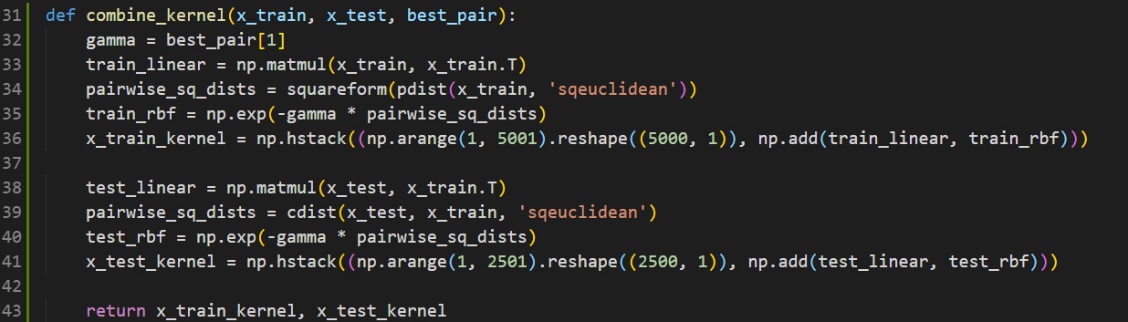


Above is grid search’s results.As we observed, best parameter (C, γ) = (16, 0.03125) and best cross validation accuracy is 98.64%.After finding best parameter, select C and γ as training parameters.We can observe that after grid search accuracy can reach 98.52%, better than 2-1’s best result.

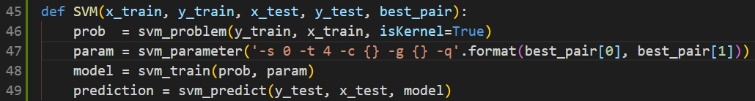
2-3.



First we read\_csv to get training data and testing data.I use best parameters which is got from previous part as our training parameter.In line 54th, through function called combine\_kernel to calculate our custom kernel function.In line 55th is to use our kernel result to train and predict our testing data.



In lines 33rd and 38th, first calculate linear kernel, also known as their inner product.Lines 34th and 39th use pdist, cdist and squareform from scipy.spatial.distance to calculate their distances.Line 39th use cdist is because we are dealing testing data not only training data, so use cdist to calculate distance from different collections.Because we need to also store data’s serial number, in line 36th and 41st, use numpy arrange to get index number from 1 to 5000 (1 to 2500 in testing data). Finally, use numpy.add to add linear kernel and rbf kernel.Last but not least, use numpy.hstack to store index number we generated into our kernel’s result.



Function SVM is actually quite same as part2-2 function called TestBest.There are two differences.One is because we use own defined kernel, we need to set isKernel=True in line 46th,the other is to set kernel\_type precomputed kernel, this is why parameter “-t” is 4.



As we can see, accuracy of mixed kernel function is 95.64%, which is a little lower part2-2, but this result is a little better than part 2-1.

So we can deduced that best result is after tuning all parameters by using grid search, while same parameter won’t cause higher accuracy by mixing two kernel functions.

1. Reference

[1] [A Practical Guide to Support Vector Classification.](https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf?fbclid=IwAR3oFHEMkb_lQgTNJXVReFxKJMhbBDVCs1gJfWxFF2wkHPb7kVLmyKD-T0w)

[2] [Libsvm for Python](https://www.itread01.com/content/1496679627.html?fbclid=IwAR1I4Sw5iIQ1c5S2WLHSHGukO-b0I594wuVPjuabvDHvuJU59ZiV96VJJyc)

[3] [scipy.spatial.distance.pdist document](https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.distance.pdist.html)

[4] [scipy.spatial.distance.cdist document](https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.distance.cdist.html)

[5] [rational quadratic kernel](https://www.cs.toronto.edu/~duvenaud/cookbook/)