Assignment 8

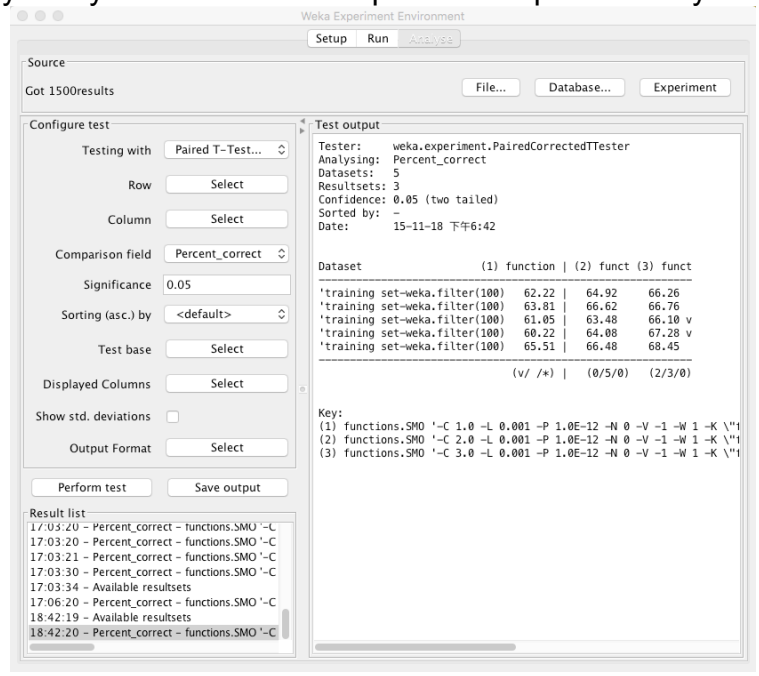
Name: Jingyuan Liu, AndrewId: jingyual

1. Complete the week 10 and 11 assigned readings about linear and statistical models, and review the lecture slides related to doing optimization in the Experimenter.

2. Use the experimenter to find the optimal setting for the Exponent parameter of the polynomial kernel on the SMO algorithm. Try 1.0, 2.0, and 3.0 as potential settings for that parameter. Evaluate the results in terms of Kappa. Fill in the following table (except for the Test set performance) from the results you see in the Experimenter:

Fold	Train Set Performance For Exponent = 1.0	Train Set Performance For Exponent = 2.0	Train Set Performance For Exponent = 3.0	Optimal Setting	Test Set Performance
1	0.6222	0.6492	0.6626	3.0	0.714268
2	0.6381	0.6662	0.6676	3.0	0.685714
3	0.6105	0.6348	0.6610	3.0	0.710145
4	0.6022	0.6408	0.6728	3.0	0.666667
5	0.6551	0.6628	0.6845	3.0	0.685714

Also take a screen shot of the Experimenter when you do the analysis so that we can verify that you are able to interpret the output correctly.



3. Now, using the Explorer and the optimal setting you observed on each fold, compute the test set performance by training a model using the optimal setting over the training data on that fold and then applying that model to the test data for that fold. Fill in these values under Test Set Performance.

4. Now using the Explorer and the whole set of data rather than the separate folds, determine what setting you would use to build a model over the whole set. What is that setting, and how would you estimate what performance you would get for that model over a new set of data?

I would use the setting of:

SMO '-C 3.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1
-K functions.supportVector.PolyKernel -C 250007 -E 3.0

To estimate the performances for this model in this dataset, we could use 10 fold cross validation.

To estimate the performances for a new dataset, we would do it like in these steps. First of all, we could split those datasets into several holds and then found out the best parameters. Then test on the held out test datasets.

5. Looking at the comparison across settings and thinking about what the different settings mean in terms of what the algorithm is doing with the data when it builds its model, what conclusions can you draw?

The different settings of the kernel would transfer the dataset to different spaces. For example, if we use different kernels like Gaussian Kernel or Poly Kernel, the transformation of data would be different

On the other hand, setting different C would cause the change of support vector numbers.

So to conclude, we could try the setting on splits of dataset to find the best settings of an alogorithms.