1. **Mini-project: Internet Advertisement, assigned today, due Oct 3rd**
2. Given three models you have built, do the following operations:
3. Pick 1 or 2 parameters of the model, construct Tables like this. Replace parameter1 with the parameter name and value1 with the real number that you use

|  |  |  |  |
| --- | --- | --- | --- |
| Parameters\Model | Gaussian Process | Random Forest | AdaBoost |
| Parameter1: value1 |  |  |  |
| Parameter1: value2 |  |  |  |
| Parameter1:value3 |  |  |  |
| Parameter2:value1 |  |  |  |
| Parameter2:value2 |  |  |  |
| Parameter2:value3 |  |  |  |

1. Tune the parameters as you did in exercise 2 of last section. Fill in the blanks with training errors.

|  |  |
| --- | --- |
| Parameters\Model | Random Forest |
| NumFeature: 1 | 0.0815 |
| NumFeature: 10 | 0.0768 |
| NumFeature: 100 | 0.0635 |
| NumTrees:2 | 0.0926 |
| NumTrees:20 | 0.0771 |
| NumTrees:200 | 0.0799 |
| Parameters\Model | AdaBoostM1 |
| NumIteration: 10 | 0.1759 |
| NumFeature: 50 | 0.1196 |
| NumFeature: 100 | 0.0981 |
| Classifier: Random Forest | 0.0356 |
| Classifier: Naïve Bayes | 0.0459 |
| Classifier: Random Tree | 0.0559 |
| Parameters\Model | RBF |
| NumCluster: 2 | 0.4243 |
| NumCluster: 20 | 0.3249 |
| NumCluster: 40 | 0.2384 |
| Ridge: 1e-10 | 0.4243 |
| Ridge: 1e-6 | 0.4243 |
| Ridge: 1e-4 | 0.4243 |

1. For each model you have built, tune the parameters to get the minimum value of training error. Write them into a table.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Random Forest | AdaBoostM1 | RBF |
| Best Parameter Set | NumFeature : 300, numTrees:1, seed:10 | Classifier: random forest, numIteration:20, resampling:false,weightThreshold:100 | Clustering seed: 1,minStdDev:0.2, numCluster:40, ridge: 1E-8 |
| Minimum training error | 0.0441 | 0.028 | 0.2368 |

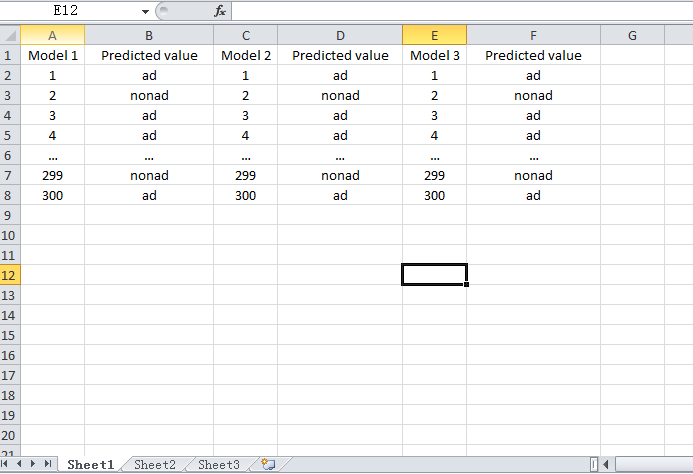
1. Follow the steps in section I to make a prediction. The test dataset “data\_prep\_300\_test.csv” is given. There are 300 instances, 1558 attributes, without labels.
2. Save your result into .CSV format, and then extract only the “**Predicted value**” column from it. I prefer the final file you sent to me is in .csv and it looks like in Fig. 14 shown below.  
     
   

Fig. 14 Template for file submission

You need to send me two things this time. The first two tasks in one document, and the third task in another document.

1. The completed table in task one. You must specify the parameter set in the table, like (M = 1, R=2, S = 10) instead of writing Parameter set 1.
2. The minimum training error table in task two (1).
3. The prediction results in .csv or.xls format, as the template shown in Fig. 14.