CISC 5352 Financial Data Analytics Quiz (6)

Kernel map approximation and selective learning (50 points)

1. Use SVM to predict implied volatility for two datasets: cisc5352.quiz6.option.csv and NBoption.csv by using at least four different kernels (You need to do your own data partition)
2. Compare its running time and running results with those of original SVM, Gradient Boost, Random Forests (The classification measure should include diagnostic index)
3. To find best performance by learning to tuning parameters for Gradient Boost,
4. Do kernel map approximation by using RBF kernel and compare its running time and running results with SVM
5. Apply selective learning to the two datasets by choosing the Gradient Boost as the learning machine.
6. *Note: you need to use plots to support all your results/conclusions*

Baby PCA (10 points)

Write a program to find PCs for the following data to finish the following tasks

$$X = \begin{pmatrix} 1 & 2 & 0 \\ 7.2 & 5 & 9 \\ -3 & 100 & 5.8 \\ 1 & -90 & 9.7 \\ 2 & 88 & 1.2 \end{pmatrix}$$

- 1. Compute its covariance matrix (unbiased)
- 2. Compute its PCs, variances
- 3. Verify the PC matrix U is an orthogonal matrix: $U^t U = I$
- 4. Compute its newdata
- 5. Retrieve the original data from the new data.

PCA Applications (40 points)

- Write a python program called demoPCA_p.py to conduct the same analysis as we did in R-version PCA analysis demoPCA.R and apply it to the two datasets. (30 points).
  - Use Pandas in your implementation
  - Also show the 3-PC visualization in addition to biplot (2-PC visualization)
- Suppose k PCs are selected in your PCA for the SP 2010_baby.csv data. It means i^{th} stock will have coordinates $p_{i1}, p_{i2} \cdots p_{ik}$ in the subspace spanned by PCs.
  - 1. find the top 20 stocks with largest PC_1 values
  - 2. find the top 20 stocks with largest PC_2 values
  - 3. Find the top 20 stocks with largest PC_k values
  - 4. Rank each stock by their PCA ranking score $\varrho = \sqrt{p_{i1}^2 + p_{i2}^2 + \cdots + p_{ik}^2}$, list the top-ranked 20 stocks, and describe their characteristics.
  - 5. Discuss the relationships between the variables according to your PCA analysis

What should you turn in?

- 1. A folder that contains
  - A report to show details of your analytics (at least 15 pages)
  - your data
  - source files
  - corresponding related output.
- 2. Please name your folder last_name1_last-name2_CISC5352_Quiz_6. For example,
- 3. Send the zipped file (.zip instead of ,rar) of your folder to Blackboard before 11:59 pm Dec 12, 2017