# Machine Learning Final Exam

**Due on 1/13/2016, 5:00pm**

**Announcement:** Please follow the guidelines for writing your report:

- The presentation should be as formal as possible (with a title page, names and IDs, and necessary outputs and pictures);
- Describe the basic ideas of each method you use and interpret clearly the results;
- Do not attach any R code in your report;
- Writing either in English or Chinese is fine.
- Please drop your report into my mailbox at the Statistics Dept office before the deadline.

The Air Pollution data set (air_pollution.txt) was collected from 41 US cities comprised with 8 variables (PT, CO, SO2, PSI, Temp, Man, Pop, Rain). Answer Q1-Q4.

**Q1**: Perform a complete Principal Components Analysis and carefully interpret the result.

**Q2:** Perform an Exploratory Factor Analysis for this data set and compare the result with the PCA solution obtained in Q1.

**Q3:** Perform a Multidimensional Scaling Analysis for this data set, do there exist clustering patterns among the cities?

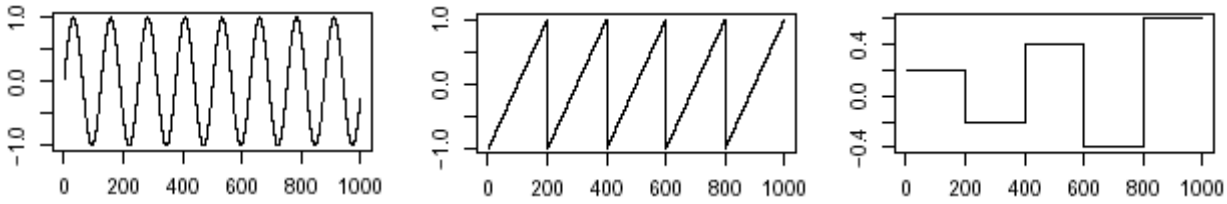
**Q4:** Suppose a researcher is interested in finding how the density of air pollution contents (PT, CO, SO2) affected by the set of variables (Temp, Man, Pop, Rain). Perform a complete Canonical Correlation Analysis for these two groups of variables and interpret the result.

**Q5**: The "wages.txt" data set contains information regarding wages and other characteristics from a random sample of 534 persons. Perform a Multiple Correspondence Analysis to identify patterns in this data set and comment on the results.

**Q6:** The Taiwanese government was informed by FBI for a possible ISIS terrorist attack on Xmas eve near Taipei 101. The columns of the data set "mixing.txt" correspond to 3 mixed dialogs (with noise) of ISIS members intercepted by FBI.

a) Please help FBI recover the original voice signals by using Independent Component Analysis.

b) The following pictures are the true dialogs of these 3 ISIS members, do your recovered signals agree with them?




**(Grading Policy: A+ = 95, A = 90, A- = 85, B+ = 80, B = 75, B- = 70, C+ = 65)**