EECS 940 Theoretic Foundation of Data Science

Instructor:

Name: Dr. Luke Huan  
Meeting Hours: Wednesday 8:30-11:00 at ITTC 250  
Office Hour: Wednesday 1:00-2:00 Eaton 2134  
Phone: 864-5072  
Email: jhuan@ku.edu  
Class Web Page: http://people.eecs.ku.edu/~jhuan/EECS940_S14

Class Objectives:

We will review statistical and mathematical principles that are utilized in machine learning and data mining research. Covered topics include asymptotic analysis of parameter estimation, sufficient statistics, model selection, information geometry, function approximation and Hilbert spaces.

Prerequisite: EECS 738, EECS 837, EECS 844 or equivalent.

Text Book:

Recommended (not required):

Grading:

<table>
<thead>
<tr>
<th>Assignment Type</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework assignments</td>
<td>25pts</td>
</tr>
<tr>
<td>One in-class presentation</td>
<td>25pts</td>
</tr>
<tr>
<td>Final Project</td>
<td>25pts</td>
</tr>
<tr>
<td>In-class discussions</td>
<td>25pts</td>
</tr>
<tr>
<td>Total</td>
<td>100pts</td>
</tr>
</tbody>
</table>
We will use the following scale to assign final grades (tentative and curving will be used):

<table>
<thead>
<tr>
<th>Grade</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>over 90%</td>
</tr>
<tr>
<td>B</td>
<td>80% - 89%</td>
</tr>
<tr>
<td>C</td>
<td>70% - 79%</td>
</tr>
<tr>
<td>D</td>
<td>60% - 69%</td>
</tr>
<tr>
<td>F</td>
<td>below 60%</td>
</tr>
</tbody>
</table>

**Extra Credit:**

Extra credits will be given to creativity and/or additional efforts shown in the team project and exams. Details will be given in the related assignments.

**Attendance:**

I expect you to come to lectures on a regular basis and will generally be unwilling to answer questions about material covered in a class you missed (unless you were sick or had another legitimate excuse). You are responsible for all announcements made in class. Participation is encouraged; please feel free to stop me if you do not understand something that has been said.

**Academic Misconduct:**

The department, school and university have very strict guidelines regarding academic misconduct. Obviously, copying is not allowed on exams. Students are expected to submit their own work on individual programming projects. Lending or borrowing all or part of a program from another student is not allowed. Students ARE allowed to borrow and modify any code on this class web site in their labs or programming projects. Instances of cheating will result in a loss on one letter grade in the course and referral to the department chairman and the dean of engineering. If a second case of academic misconduct is reported in any class, a dismissal hearing may be initiated by the dean of engineering.

**Topics Covered (subject to change during the course):**

1. Measure, $\sigma$-algebra, random variables
2. Law of large numbers
3. Statistical decision theory, subjective Bayesian
4. Maximum likelihood, Bayesian, Minimax
5. Linear regression
6. Linear classification
7. Bayesian regression
8. SVM
9. Decision tree, Boosting, Boosted Decision Tree
10. Gaussian processes
11. EM
12. Statistical Learning Theory
13. Approximate inference in Bayesian
14. Probabilistic graphical models
15. Bayesian nonparametrics
16. Big data
17. Other topics related to Machine learning and data mining