POLITECNICO DI MILANO

Dipartimento di
Elettronica e Informazione

Pattern Analysis and Machine Intelligence

***M. Matteucci**, Luigi Malagò, Davide Eynard*

matteucci@elet.polimi.it
Dipartimento di Elettronica e Informazione, Politecnico di Milano
Artificial Intelligence and Robotics Lab

Pattern Analysis and Machine Intelligence

A new course … more or less

- Originally named as Methodologies for Intelligent Systems
- New name, new program, new teachers (more or less …)

Lectures given by:

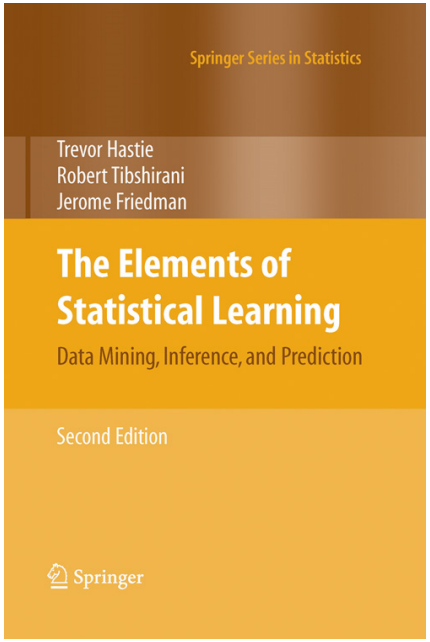
- Matteo Matteucci (Lecturer)
  http://www.elet.polimi.it/people/matteucc
  +39 02 2399 3470
  matteucci@elet.polimi.it
- Luigi Malagò (Teaching Assistant)
  http://www.elet.polimi.it/people/malago
  malago@elet.polimi.it
- Davide Eynard (Teaching Assistant)
  http://www.elet.polimi.it/people/eynard
  davide.eynard@polimi.it

Course website: https://chrome.ws.dei.polimi.it (or from my dei web page)

Material & Schedule

Course Material on Machine Learning (tentative)

- The Elements of Statistical Learning, T. Hastie, R. Tibshirani, J.Friedman, Springer, 2008
  - Great book!
  - Available at the Polimi library
  - Available online as pdf on the book website!
- Additional material on the course website so keep checking the course website
- Only few slides available, you need to take notes …



Course schedule

- Tuesday from 13:15 to 15:15 room 3.6 (lecturer)
- Wednesday from 14:15 to 16:15 room 4.2 (teaching assistants)

More Important Stuff …

The goal(s) of the course

- Introduce most used techniques from ***Machine Learning*** and ***Pattern Recognition***
- Provide a ***sound (statistical) background*** to understand them
- Introduce ***Data Mining*** basics

Evaluation and Grading !!

- Oral (or written?) final examination with theory questions and simple exercises [***70% of grading***].
- Mandatory homework (due in 15 days) to integrate the theory (this can be substitued with a project to be discussed in advance with the teacher) [***30% of grading***].

Detailed schedule and topics …

- Published on the website as soon as we decide for the class days
- Selected chapters from the book and additional material

A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence

J. McCarthy, Dartmouth College M.L. Minsky, Harvard University
N. Rochester, I.B.M. Corporation C.E. Shannon, Bell Telephone Laboratories
August 31, 1955

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

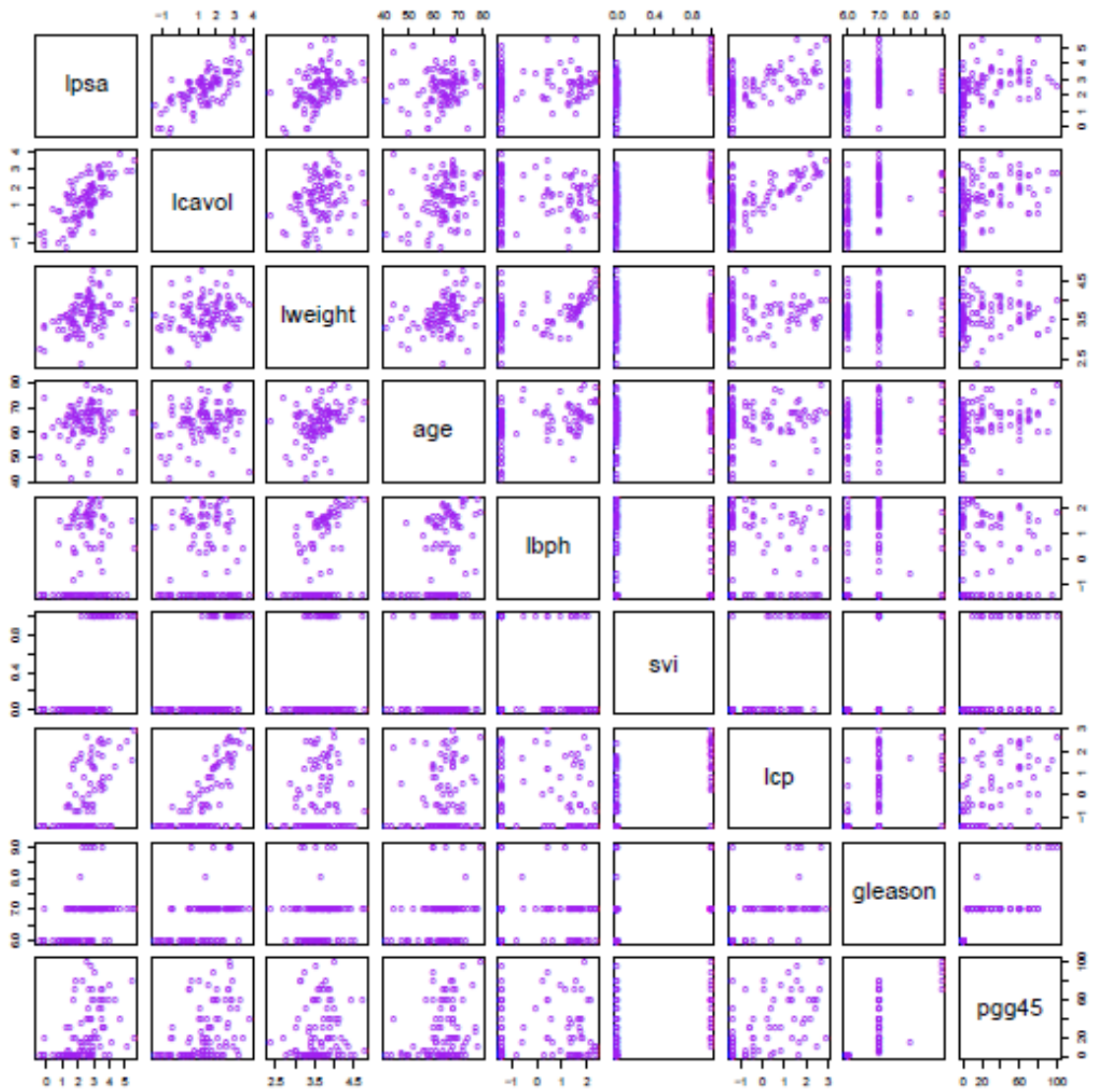
http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

Email SPAM

TABLE 1.1. *Average percentage of words or characters in an email message equal to the indicated word or character. We have chosen the words and characters showing the largest difference between* `spam` *and* `email`.

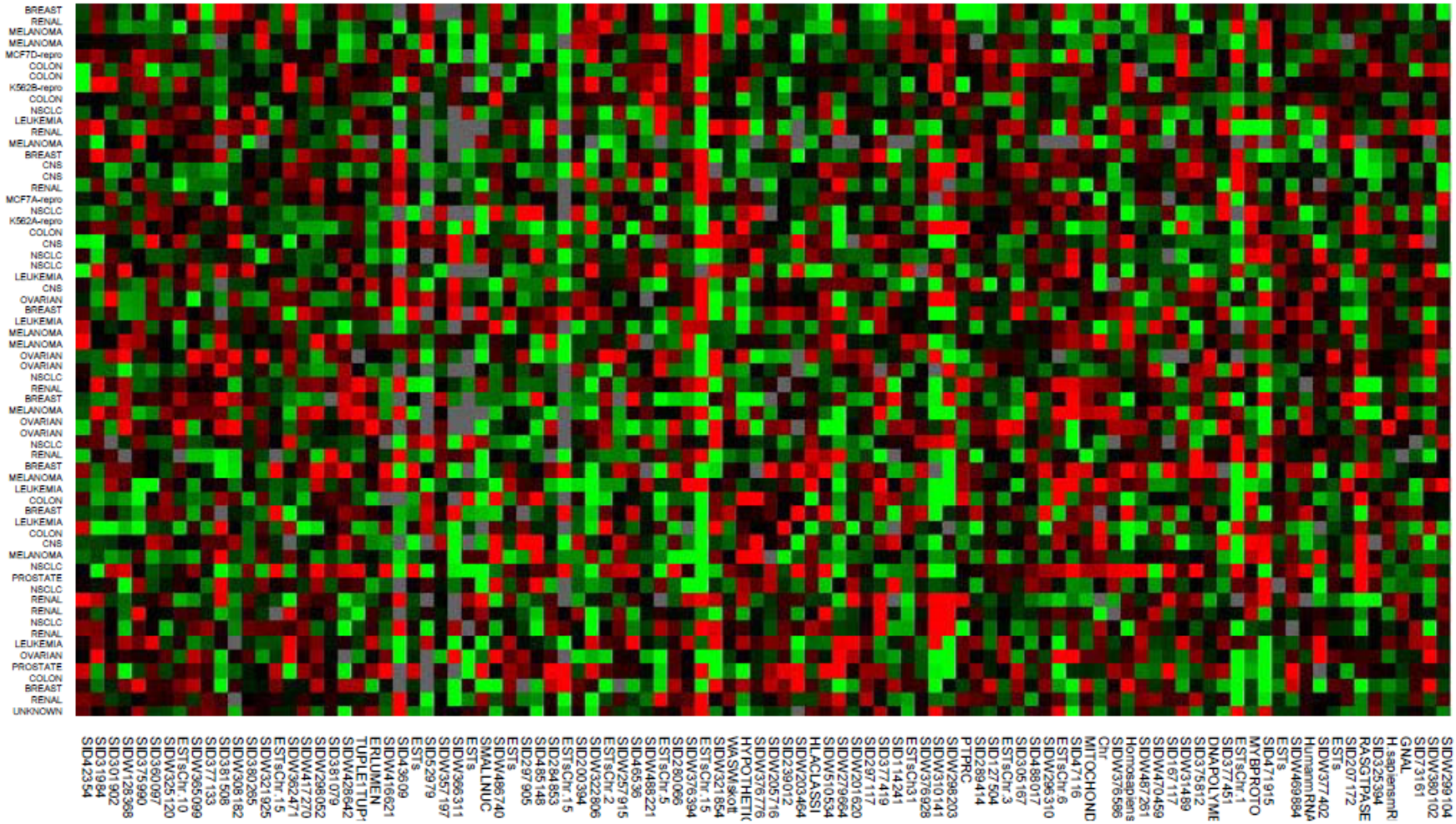
	george	you	your	hp	free	hpl	!	our	re	edu	remove
spam	0.00	2.26	1.38	0.02	0.52	0.01	0.51	0.51	0.13	0.01	0.28
email	1.27	1.27	0.44	0.90	0.07	0.43	0.11	0.18	0.42	0.29	0.01

Prostate Cancer



Handwritten Digit Recognition

DNA Expression Microarrays



Learning Paradigms

Imagine an organism/machine that experiences a series of sensory input:

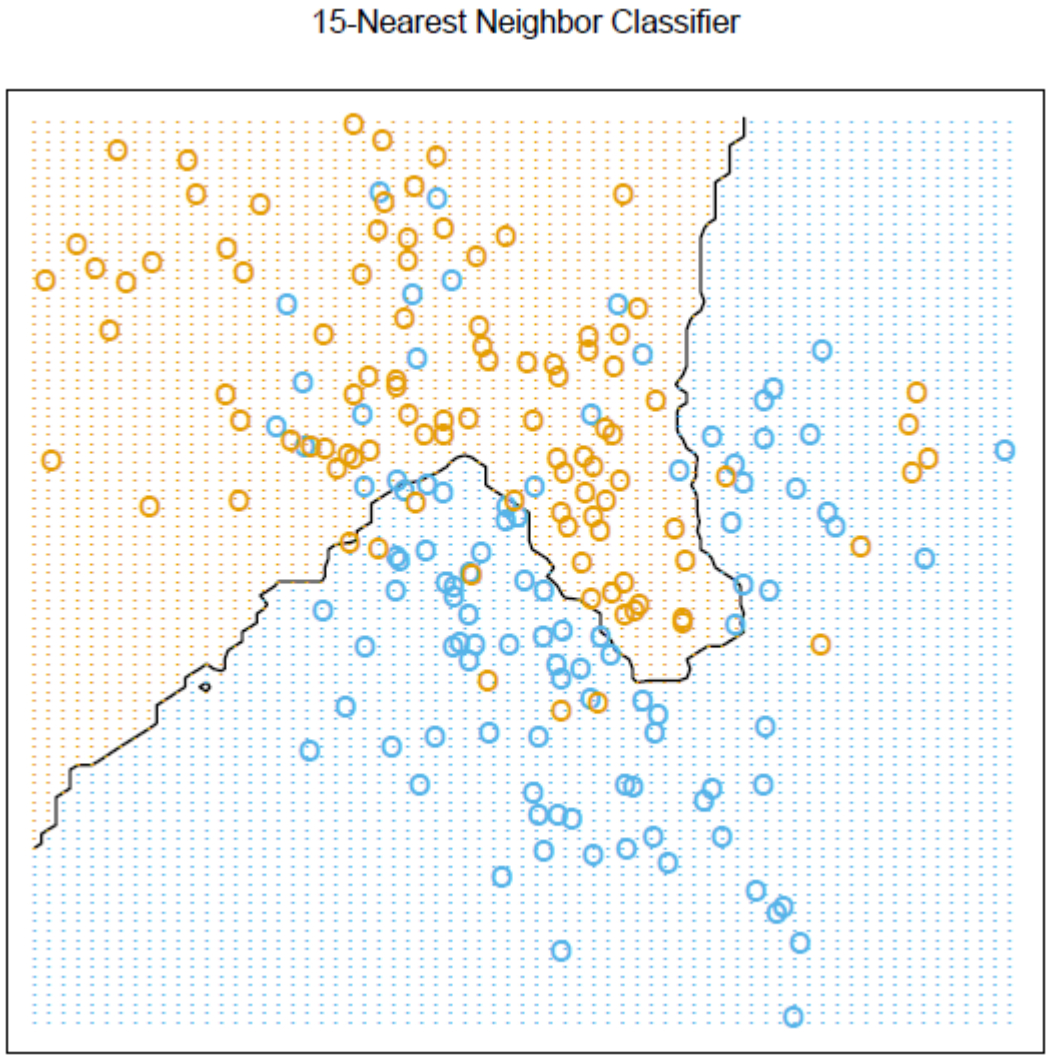
$$E = x_1, x_2, x_3, x_4, \ldots$$

- Supervised learning: given the desired outputs $y_1, y_2, \ldots$, learn toproduce the correct output given new input
- Unsupervised learning: exploit regularities in E to build a representation that can be used for reasoning or prediction
- Reinforcement learning: producing actions $a_1, a_2, \ldots$ which affectthe environment, and receiving rewards $r_1, r_2, \ldots$ learn to act in a way that maximizes rewards in the long term

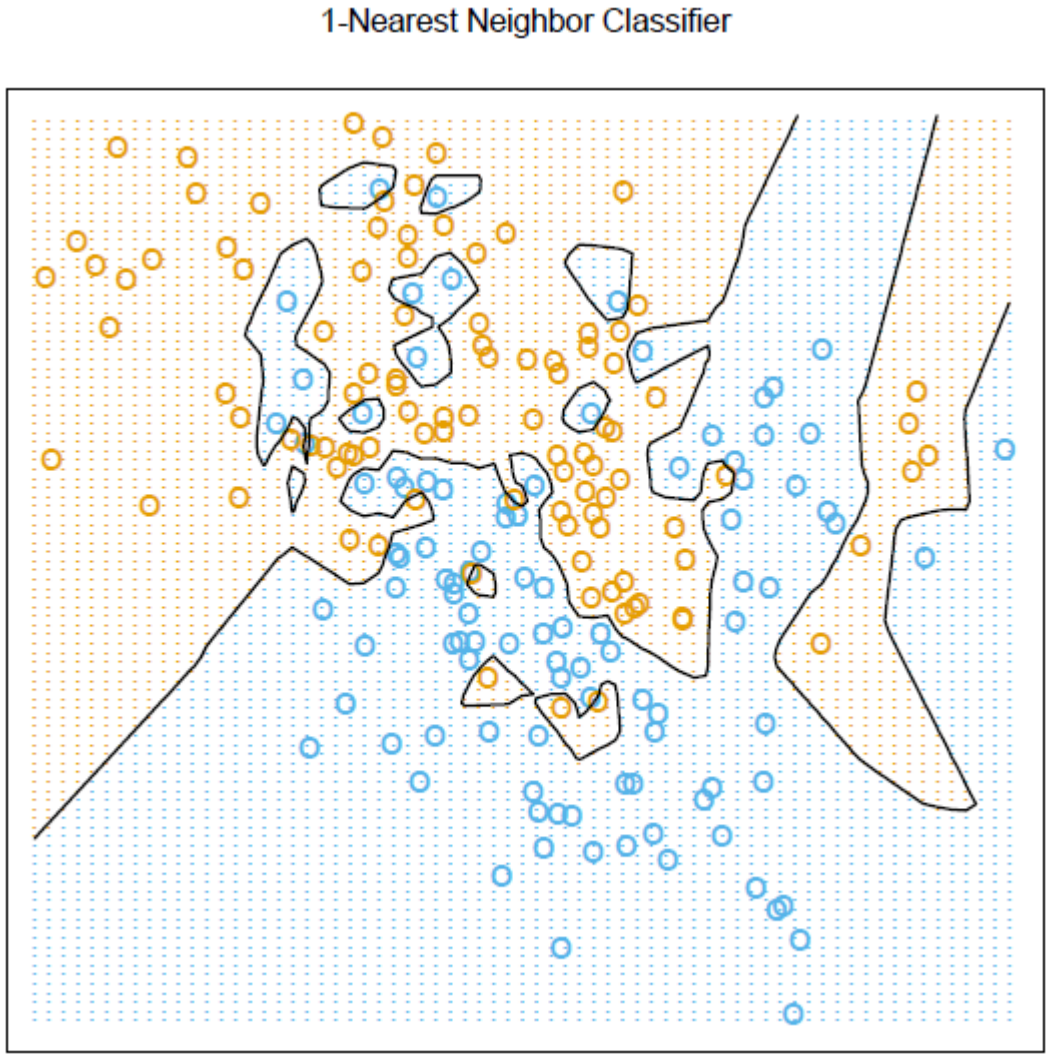
Course focus on Supervised Learning,
but you'll some Unsupervised Learning as well

Linear Regression

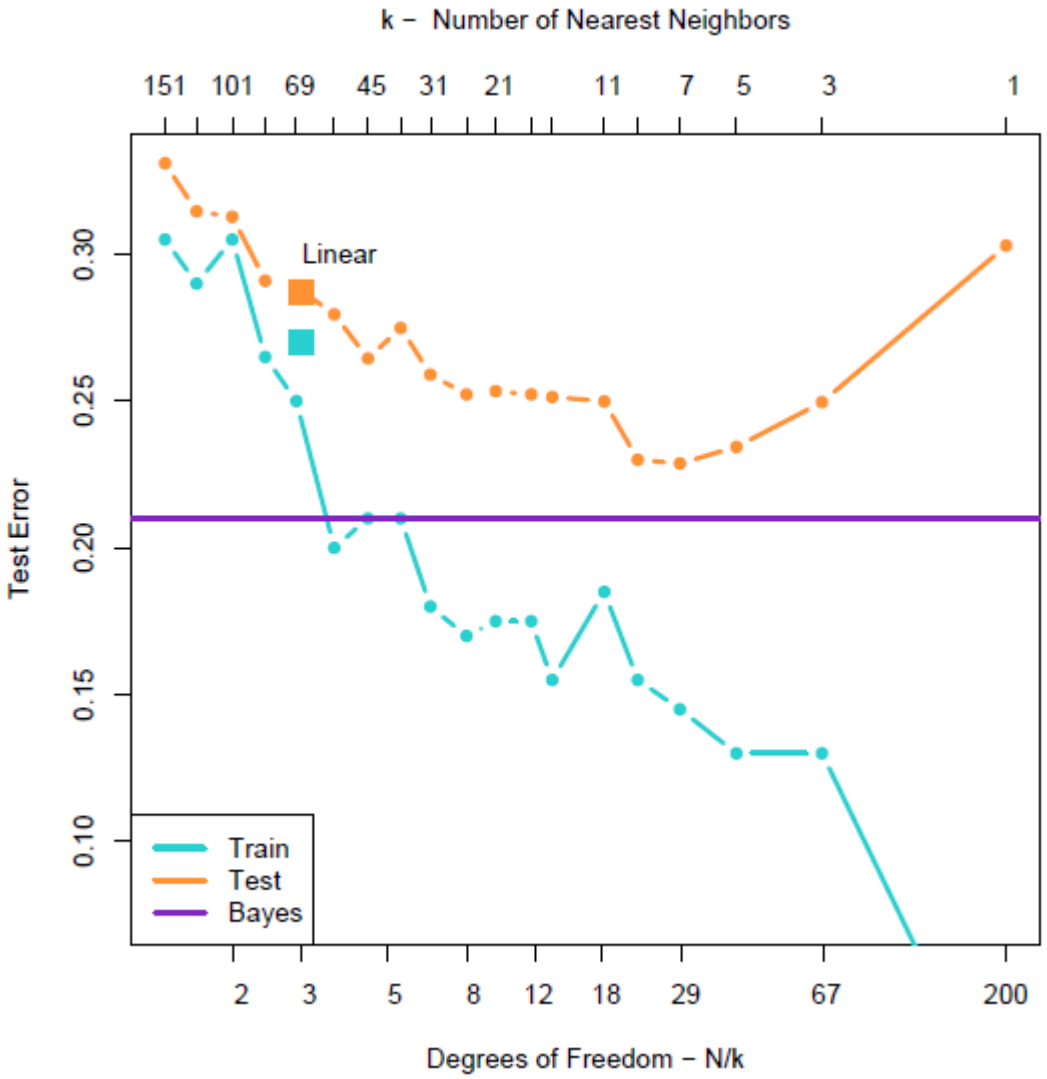
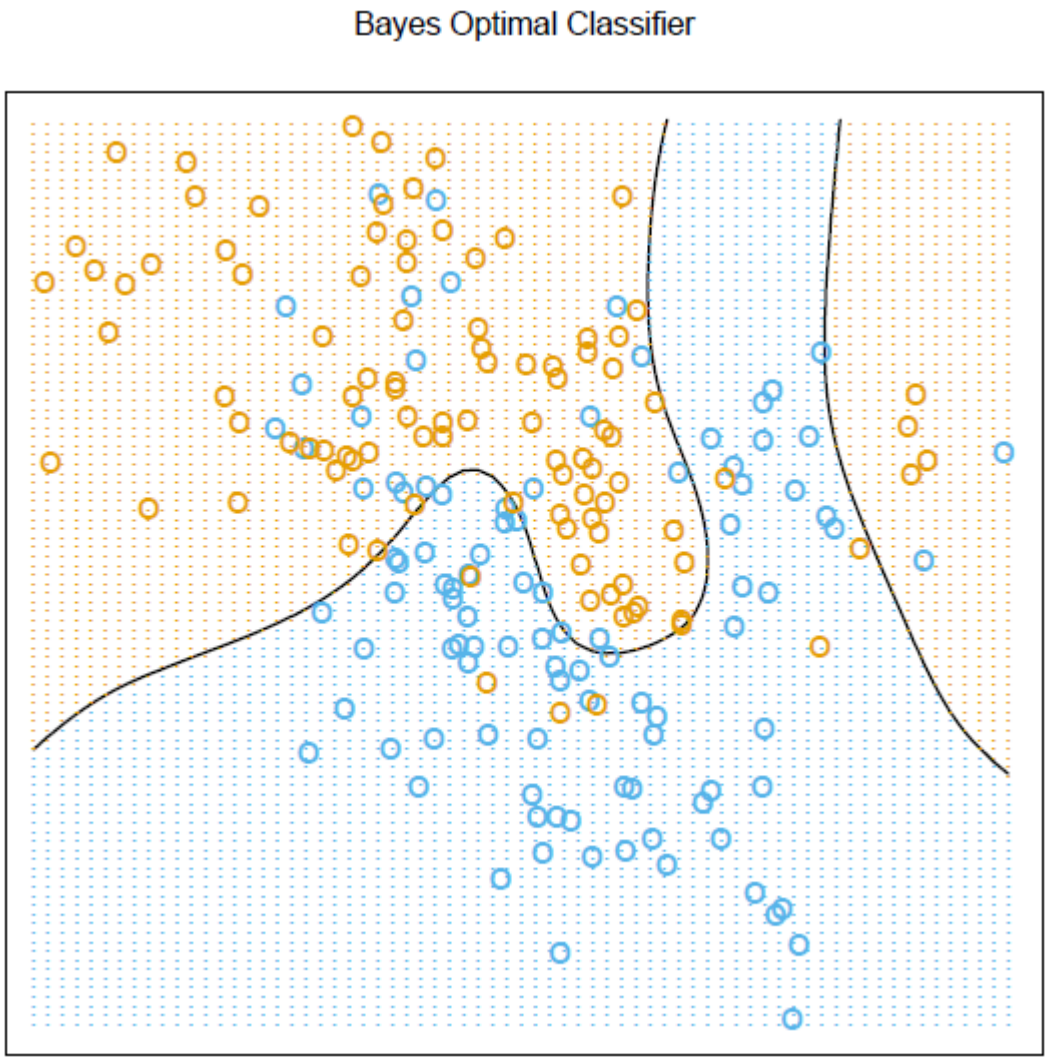
K - Nearest Neighbor Method (K = 15)



K - Nearest Neighbor Method (K = 1)



Misclassification Curve (Test Set Error)



Optimal Bayes Classifier