EOSC 510 Data Analysis in Atmospheric, Earth & Ocean Sciences

William W. Hsieh
Department of Earth, Ocean & Atmospheric Sciences
The University of British Columbia

Course web site:
http://www.ocgy.ubc.ca/~william/EOSC510/

The Big Picture

- Course objective: Learn from data
  - Discover patterns or signals
  - Predictions
- Statistics: originated from probability theory, born when mathematicians (Pascal, Fermat, Gauss) became interested in gambling problems.
- Machine learning: a main branch of artificial intelligence (computational intelligence)
- Machine learning more interested in complexity and nonlinear behaviour, whereas statistics more interested in probability.

- Problems types:
  - (a) outcome is discrete, e.g. classification, clustering
  - (b) outcome is continuous, e.g. regression, principal component analysis
  - (c) outcome is a probability distribution.
- Machine learning most involved with type (a), and Statistics with type (c).
- Data mining: aims to discover patterns from large datasets using techniques from both machine learning and statistics.

Course website: http://www.ocgy.ubc.ca/~william/EOSC510/

Computing in this course

- Some exercises involve programming.
- A variety of programming languages can be used:
  Matlab/Octave, R, Python, … [but don't use Excel]
- I use Matlab, which has an online interactive tutorial for beginners:
  http://www.mathworks.com/academia/student_center/tutorials/launchpad.html