Linear Regression Models
W4315

Instructor: Dr. Frank Wood

Text: Applied Linear Regression Models
Authors: Kutner, Nachtsheim, Neter

Not Registered Yet?

- **Fill out the form at**
  - **http://tinyurl.com/mqfq95**

Course Description

- Theory and practice of regression analysis, Simple and multiple regression, including testing, estimation, and confidence procedures, modeling, regression diagnostics and plots, polynomial regression, colinearity and confounding, model selection, geometry of least squares. *Extensive use of the computer to analyse data*.

Course Outline

- Roughly 1 chapter per week
- 3-5 weeks, linear regression
  - Least squares
  - Maximum likelihood, normal model
  - Tests / inferences
  - ANOVA
  - Diagnostics
  - Remedial Measures
  - Linear algebra review
  - Matrix approach to linear regression

Course Outline Continued

- 3-4 weeks multiple regression
  - Multiple predictor variables
  - Diagnostics
  - Tests

  Midterm

Course Outline Continued

- 3-4 weeks on generalized regression
  - Polynomial regression
  - Logistic regression
  - Neural networks
  - Generalized linear models
- 3-4 weeks on Bayesian regression
  - MCMC
  - Bayesian linear regression
  - Gaussian process regression
  - Projects

Requirements

- Calculus
  - Derivatives
- Linear algebra
  - Matrix notation, inversion, eigenvectors, eigenvalues
- Probability
  - Random variables
  - Bayes Rule
- Statistics
  - Expectation, variance
  - Estimation
  - Bias/Variance
  - Basic probability distributions
- Programming

Projects (homework and final)

- Software
  - Don’t care.
    - R
    - Matlab
    - S-Plus
    - SAS
    - Minitab
    - Excel
    - java, c++, c, assembly, …

Grading

- Bi-weekly homework (35%)
  - Due every other week
    - no late homework accepted
  - One allowed to be missed
  - Completing all is “extra-credit”
- Participation (5%)
- Midterm examination (25%)
- Final project (35%)
- Curve

Office Hours / Website

- http://www.stat.columbia.edu/~fwood
- Office hours : Tuesday 1-3pm
- Office : Room 1011
- TA : Heng +.5 Xiaoru
  - TA office hours TBD

Why regression?

- Want to model a functional relationship between an “predictor variable” (input, independent variable, etc.) and a “response variable” (output, dependent variable, etc.)
  - Examples?
- But real world is noisy
  - Observation noise
  - Process noise

History

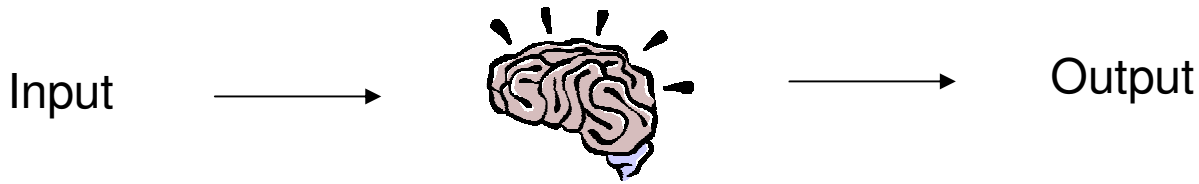
- Sir Francis Galton, 19th century
  - Studied the relation between heights of parents and children and noted that the children "regressed" to the population mean
- "Regression" stuck as the term to describe statistical relations between variables

Example Applications

- Trend lines, eg. Google 6 last 6 mo.

Others

- Epidemiology
  - Relating lifespan to obesity or smoking habits
  - etc.
- Science and engineering
  - Relating phsyical inputs to physical outputs in complex systems
- Grander



$$\text{output} \sim f(\text{input})$$