Machine Learning applied to Planetary Sciences

PTYS 595B/495B

Leon Palafox

https://leonpalafox.github.io/MLClass/

News of the day

Today, $900 million is being awarded from the Fund for the following 13 initiatives selected through a rigorous merit review process:

Lead Institution	Initiative Title	Amount of Funding over Seven Years
Dalhousie University	Safe and Sustainable Development of the Ocean Frontier	$93,732,000
Laurentian University	Metal Earth	$49,269,000
McGill University	Healthy Brains for Healthy Lives (HBHL)	$84,000,000
Polytechnique Montréal	The "Montreal TransMedTech (MT2) Institute": An open-innovation ecosystem for the development of next generation medical technologies	$35,625,000
Queen's University	Canadian Particle Astrophysics Research Centre	$63,744,000
Université de Montréal	Data Serving Canadians: Deep Learning and Optimization for the Knowledge Revolution	$93,562,000
University of Alberta	Future Energy Systems Research Institute	$75,000,000
University of Calgary	Global Research Initiative in Sustainable Low Carbon Unconventional Resources	$75,000,000
University of Guelph	Food from Thought: Agricultural Systems for a Healthy Planet	$76,613,000
University of Saskatchewan	World Water Futures: Solutions to Water Threats in an Era of Global Change	$77,840,000
University of Waterloo	Transformative Quantum Technologies	$76,277,000
Western University	BrainsCAN: Brain Health for Life	$66,000,000
York University	Vision: Science to Applications (VISTA)	$33,338,000
	TOTAL	**$900,000,000**

http://www.marketwired.com/press-release/minister-bibeau-announces-213187000-transform-research-universite-de-montreal-hec-montreal-2155797.htm

Up until now

- Machine Learning
  - Discipline to help do automated tasks
- Machine Learning is a sub-discipline of Math
  - We need linear algebra
- In Machine Learning we often deal with uncertain scenarios.
  - We need probability
- We have two kinds of ML:
  - Supervised
  - Unsupervised

First Tool - Linear Regression

- Imagine you want to sell you car:
  - How much do you ask for it:
    - Mileage
    - Year
    - Color
    - Options
    - Condition

What do places like Edmunds.com do?

Model	Year	Brand	Price	Options	Condition	Mileage
Corvette	1961	Chevrolette	100K	Standard	As New	100,000
Corvette	1961	Chevrolette	10K	Standard	Rust	100,000
Corvette	1961	Chevrolette	120K	Standard	Used	20,000

Price = Year + Options + Condition + Mileage

Price = A*Price + B*Options + C*Condition+ D*Mileage

Linear Regression

- We are trying to predict a variable in the continuum.
- We have features or independent variables.
- We want to find the best set of weights to generalize the regression.

Other applications for regressions

- House Prices
- Budget for a Movie
- Effect of treatment
- Political inclination (1-100)
- Number of Likes
- Number of sales (for a product)

Formally

$$y_1 = \beta_1 x_{11} + \cdots + \beta_p x_{1p}$$
$$y_2 = \beta_1 x_{21} + \cdots + \beta_p x_{2p}$$
$$y_3 = \beta_1 x_{31} + \cdots + \beta_p x_{3p}$$

$$y_n = \beta_1 x_{n1} + \cdots + \beta_p x_{np}$$

Matrix form

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \mathbf{x}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix},$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} \qquad \mathbf{y} = \mathbf{X}\beta$$

Optimization problem

- Optimize:

$$\mathbf{y} = \mathbf{X}\beta$$

- Score:

$$X\beta - y$$

$$(X\beta - y)^2$$

$$\frac{1}{2}(X\beta - y)^2$$

Two ways to approach the problem

- Matrix Algebra (next class)

- Gradient Descent (next class)

What happened with the features?

- Why did our original simple approach did not work
- What do you think the intercept term does?
  - We didn’t mention that in class
- What would happen if we increase the number of features?