

CDS Education We explore, learn, and educate big minds.

Updates

CMS will be up tonight.

Work on Project A (due next Tuesday Midnight).

If you joined late, permission numbers will be sent out soon.

+ Extended quiz deadline





Intro to Machine Learning and Linear Regression

Machine Learning

According to Wikipedia...

"Machine Learning is a subfield of computer science that gives computers the ability to learn without being explicitly programmed."







Better Definition

By Professor Kilian Q. Weinberger

CS 4780/5780







Traditional Computer Science









Machine Learning













My Definition

- Let's talk about functions...
- f(x)? =x? =x^2?
- Vertical line test?







ML Setup

Hypothesis: Some speculative relationship between the input space and output space

Input Space: Variable or set of variables(data)

Output Space: Target variable to estimate



Supervised vs Unsupervised

Supervised learning problems...

- Known target variable info
- Validation examples

Unsupervised learning problems...

- Unknown target variables
- Difficult to validate





Supervised learning:

given $(x_1, y_1), ..., (x_n, y_n)$, learn f(x) = y

Unsupervised learning:

given x_1, \ldots, x_n , learn patterns or structure

- Online learning: for i = 1, ..., n, given x_i, predict and observe y_i, learn f(x) = y
- Active learning: for i = 1, ..., n, choose x_i, predict and observe y_i, learn f(x) = y
- Reinforcement learning: for i = 1, ..., n, choose x_i, predict y_i, observe reward r_i, learn f(x) = y

Supervised Learning



Setup

- Training / Validation split
- Feature variable(s)
- Target variable
- Train and Test





Machine Learning









Validation Set

- Split data into two sets
- Train model on one and validate on another
- Advantages / Disadvantages?







Output Space Properties

- **Continuous -** e.g. temperature, height, probability
- **Discrete -** e.g. car brands, race, Pokémon type, diagnosis





Regression vs Classification





Classification

Regression

What is Learned





Objective function

- All ML problems are optimization problems
 - Format: Minimize/Maximize Obj in terms of x.
 - Subject to set of constraints
- Objective functions represent assumptions
- Value of objective is an estimation of error



Calculating Error





Linear Regression









Linear Regression

$$y = B_0 + B_1 x_1 + \dots + B_p x_p + \varepsilon$$

What are the assumptions?

- Linear relationship
 - B, the coefficient vector, does not depend on x
- There is an unremovable noise
- This noise is normally distributed about the line

Objective: Least Squares Error (L2)



What does this minimize?

Why this form?





Your problem set: Project A

Next week: Introduction to Classification

See you then!



