Dis Fri

Friday, October 5, 2018 11:53 AM

Today: Math Review

Sajad Darabi

A **set** is a collection of items. We use curly braces to denote sets
A = {3, 6, 9}
$\mathbb{N}$ - set of natural numbers

Cardinality |A| is the size of set A

The universal set Ω is the set that includes all other sets in your space events. i.e. for every set A, $A \subset \Omega$

Operators:
Union ∪
Intersection ∩
Subset ⊆
Complement A^c

Properties of set operations: commutative, associative, distributive, De Morgan's

$|A \cup B| = |A| + |B| - |A \cap B|$

Determinism - events determined by previous causes
Random - cannot predict the outcome
Probability - what we use to quantify randomness

Frequentist Interpretation - run tests and approximate
Ex: rolling die. roll it many times and approximate probability of each side.
Bayesian Interpretation - use prior knowledge
Ex: rolling die again. use the fact that it has 6 sides and conclude it is probability 1/6 for each side

Bayes Rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Independence
$P(A \cap B) = P(A)P(B)$

Random Variables
Functions that map outcomes to real valued numbers
$X : \Omega \to R$

Distributions
We define a distribution of a random variable.
$x \sim p(x)$
means x was sampled from a probability distribution p(x)

Expectation
E(X) = $\sum_i P(X = a_i)a_i$
or for a continuous variable, we integrate

Variance
$Var(x) = E(X - \mu)^2$

Covariance
$cov(x, y) = E[(x - \mu_x)(y - \mu_y)]$

Calculus
We should already know:
- Derivative
- Chain Rule
- Partial Derivative

Linear Algebra
A an mxn matrix
B an nxp matrix
A and B can be multiplied, and AB is an mxp matrix

let x, y be vectors, b a scalar.
We can define:

$\frac{\partial x}{\partial b}$ a column vector

$\frac{\partial b}{\partial x}$ a row vector

$\frac{\partial x}{\partial y}$ a matrix

some books will swap the first two, just stay consistent

Problem:
$y = Ax, x \text{ and } y \text{ are column vectors size } n$
What is $\frac{\partial y}{\partial x}$?

$y_1 = A_1^T x = a_{11}x_1 + a_{12}x_2 + \quad + a_{1n}x_n$
$y_2 = A_2^T x = a_{21}x_1 + a_{22}x_2 + \quad + a_{2n}x_n$

$$\begin{bmatrix} a_{11} & a_{21} & & \\ a_{12} & a_{22} & \cdot & \cdot \\ \vdots & \vdots & & \\ a_{1n} & a_{2n} & & \end{bmatrix} = A^T$$

Problem:
$\alpha = x^T A y$
x nx1, A nxn, y nx1

$\frac{\partial \alpha}{\partial x}$?

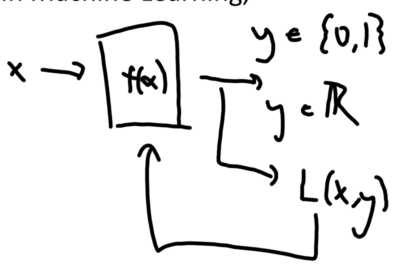
$(x_1 a_{11} + x_2 a_{21} + .)\, y_1$

$+\, (x_2 a_{11} + x_2 a_{22} + .)\, y_2$

$$\begin{bmatrix} a_{11} y_1 + a_{12} y_2 + \\ a_{21} y_2 + a_{22} y_2 + \\ \\ \end{bmatrix} = Ay$$

Eigenvalues
$Ax = \lambda x$
$Ax - \lambda x = 0$
$\det(A - \lambda I) = 0$

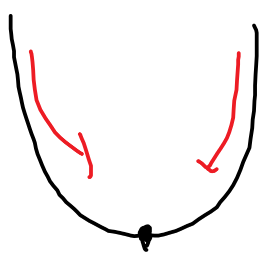
In Machine Learning,



L is a loss metric we use to train the model f(x). Generally we want to minimize L.
Ex:

$$L = \sum_{i}^{n} f(x_i) - y_i$$

How can we minimize L? If it's a nice convex function (one global minimum), we can use gradient descent:



How to tell if a function is convex? We can show that the second derivative is > 0 everywhere.