Machine Learning
8/31/2016

Probability

If everyone uses 12 sided dice P = 1/12. Each dice has the same probability. Do as many experiments as we want and characterize these statistics.

Fivethirtyeight.com: election forecast. Blue = democrat winning, red = republican; depth of color is how much that dominates.

Hilary Clinton will only run against Trump once because only one will win and unlikely to end up with same politician running again next time; won't get exact same two elections. In this case, probability doesn't represent a test you can run multiple times; it's a measure of uncertainty or certainty for an event.

Word is loaded- probability is a well-defined metric for certainty. ML deals with classifying information but we cannot make assumptions for sure.

2 kinds of probabilities:

1. **Frequentist** (can make as many experiments as you want, see how many times a certain outcome happens)
    a. Dice example
    b. Soccer matches, maybe, and NBA games- can repeat the experiment in a season but not perfect example
    c. Factory quality control (to 6 sigma, or standard deviations-99.999…% of products will come out acceptable)
    d. Is often misused-we want to use tools of frequentist even if we don't have ability to repeat an experiment… if can't do an infinite amount of experiments, shouldn't do this. "Bootstrapping" (creating many scenarios) is the tool of frequentist by excellence. Frequentist needs experiments to be done.
2. **Bayesian** (belief in a certain event happening)
    a. Other way around- no ability to do an infinite amount of experiments. So you are trying to find probability with limited amount of knowledge.
    b. Examples: plane crashes, presidential elections, medical diagnosis
    c. Represents a degree of belief in a certain event
    d. People don't use it enough

These share exactly the same rules from a math point of view; frequentist and Bayesian exactly the same in terms of math, laws, theorems. It is easy to mix them up (beware of P-values!). Want less than 0.05 (95% confidence interval) – no reason for this 0.01. Same reasoning for voltage of electrical outlets. Was used for convenience because that was needed to turn lightbulb on at the time but it's actually very arbitrary. Someone once said 0.05 is a reasonable P value. But doesn't make a difference between 0.04, 0.05 and 0.06. One of cruxes of frequentist approach.

Random Sampling
Sampling from a probability distribution- take out numbers.

TV show words in scripts:
Enterprise, Spock, Picard, Wesley (= Star Trek)
Lannister, Stark, Theon, Westeros (= Game of Thrones)
House, Doctor, Vicodin (= House)

Movies:
Clooney, Damon, Heist (=Oceans 11/12/13)
Damon, Mars, NASA (= The Martian)
Damon, Robin Williams, MIT (= Good Will Hunting)

Matt Damon appears 3 times…can have same element in different times for different probability distributions. Each movie is a prob distri and each domain is a prob distri as well. Damon appears in many movies but less TV shows.

How many probability distributions can we have? "It's turtles all the way down!"

Say you have a list of every person in the world. If we go to people related to movies; probability of Leon Palafox showing up is (a lot) less than Matt Damon. Probability distribution of people related to movies, screenwriters, actors, batman movies, etc.

Probability will never be exactly 0 so if need to divide by a probability, won't go to infinity.

If new person comes from another planet, what is chance he sees Leon vs Clooney….about the same. Alien watching tv now, chance of seeing Clooney will be higher than Leon.

Fruit examples (Bishop's) on Jupyter
(Assignment will involve Second Chapter of Yoshua Bengio's Deep Learning book)
Oranges vs Apples

Random variables (in movies, they are actors, name of movies, etc)
Boxes and fruits are random variables. 2 kinds of variables: random and deterministic
Deterministics: physical constants (speed of light, etc) not changing between experiments
Start drawing random elements: fruits and boxes, can start to get an idea of populations.
Probability of choosing red box vs blue box: say it's ½ vs ½ (for Leon's example, diff than Bishop's who uses diff probability of picking diff boxes)

Rule of probability: probabilities of all ALWAYS have to sum to 1

Probability of choosing an apple in the red box

Red box has 8 elements but 2 apples- universe convinced to red box now, is as if blue box doesn't exist anymore… 2/8

Blue box chance is ¾

When choose a box, probabilities change even though they have same elements. Could be a box doesn't even have an apple.

Given your box is red, probability of fruit being orange + apple needs to sum to 1. But can't cross probability, i.e. orange in red + orange in blue box.

Some generalizations:
(Bishop 1st tells you this part of rules of probabilities.)
Now given your universe of boxes and fruits…what is probability that you reach an apple in the red box (probably of picking out of red box AND an apple at same time)
Chain of events; probability product rule. Pick red box then pick fruit. This can be turtles all the way down….could also pick red box given that the box is 5 feet away vs 1 feet away and then add another level of probability and then probability of box being 5 feet away vs 1 feet away…can keep adding levels of probabilities. Could add probability of drawing fruits- probability function based on how close fruit is to top (may only draw fruit from top of box)

1st go into domain of box then p(box = red)
then go into how many apples within that box p(fruit=apple|box = red)
multiply… "probability product rule"

Probability that fruit = orange (don't care if came from red or blue box)- probability sum rule
P(fruit = orange) = "marginal probability" = p(fruit = orange, box = blue) + p(fruit = orange, box = red). Probability to win a game, pass a class, etc. Golden rule of probability.

Discrete (can be described by single, real number) vs continuous (have a function) probability distribution.

If event x and y are unrelated? p(x|y) = p(x) Probability of picking apple given box is red is same probability as having an apple (means doesn't matter which box you pick – happens if boxes have same distribution of fruits so doesn't matter). What is probability that Hillary Clinton wins if fly is flapping its wings in North Korea? Probably same probability as Hillary Clinton just winning. When that happens, (p(x,y) = Hillary winning given fly flap its wings) = p(x|y)p(y) = p(x)p(y) (prob Hillary winning * prob of fly flapping wings)

Bayes Rule
p(x|y) = p(y|x)p(x)/(py) is just probability of event x happening given y… will go over proof. Will be in assignment. Remember you can switch elements and calculate other marginal. This is important to start answering questions in more informed way.

Example of how to use: there was a study that found that more of a type of cancer in rural America than in cities. Economists and health professionals then started saying small town diets were bad. Town of 100 people and one person gets cancer, rate rockets to 1%. Using frequentist approach draws bad assumptions. Need to account for number of people in town…. If you assume number of people also a probability distribution, only then can you draw conclusions to see if frequentist rate in small town is same as in big city.