Motivation and Basics

Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

Weekly Objectives

- Motivate the study on
  - Machine learning, AI, Datamining….
  - Why? What?
  - Overview of the field
- Short questions and answers on a story
  - What consists of machine learning?
  - MLE
  - MAP
- Some basics
  - Probability
  - Distribution
  - And some rules…

WARMING UP
A SHORT EPISODE

Thumbtack Question

- There is a gambling site with a game of flipping a thumbtack
  - Nail is up, and you betted on nail's up you get your money in double
  - Same to the nail's down

- A billionaire wants to enter the gambling
  - With scientific and engineering supports
    - He is paying you a big chunk of money
  - He asks you
    - I have a thumbtack, if I flip it, what's the probability that it will fall with the nail's up?
  - Your response?

Number of Trials

$$\hat{\theta} = \frac{a_H}{a_H + a_T}$$

- You report your proof to the billionaire
  - From the observations of your trials, and from the MLE perspective, and by assuming the binomial distribution assumption…………..
  - $\boldsymbol{\theta}$ is 0.6
  - So, you are more likely to win a bet if you choose the ***head***
- He says okay.
  - Billionaire
    - While you were calculating, I was flipping more times.
    - It turns out that we have 30 heads and 20 tails.
    - Does this change anything?
  - Your response
    - No, nothing's changed. Same. 0.6
  - Billionaire
    - Then, I was just spending time for nothing????
- You say no
  - Your additional trials are valuable to …….

Wait!!!

A student whose name is Bayes raised his hand

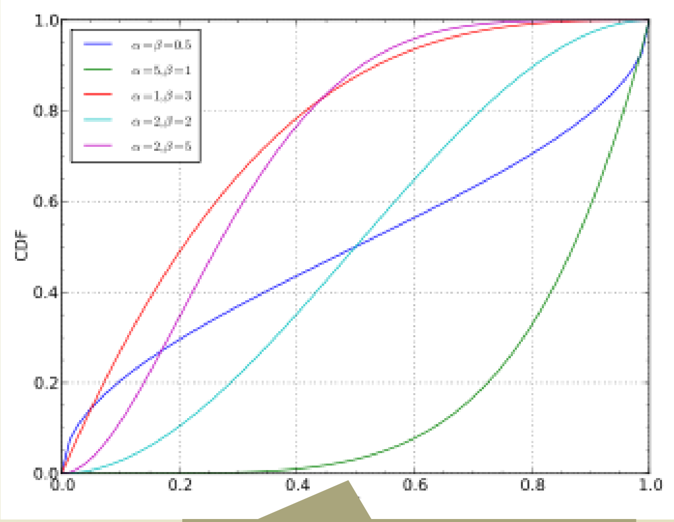
Incorporating Prior Knowledge

- Bayes says
  - Wait. Billionaire.
  - Is it really true that the thumbtack has 60% chance of head?
  - Don't you think it is 50 vs 50?
- Billionaire says
  - Well. I thought so...
  - But, how to merge the previous knowledge to my trials?
- Bayes says
  - So, I give you this theorem!
  - $P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$
  - You already dealt with $P(D|\theta) = \theta^{a_H}(1-\theta)^{a_T}$
  - P(θ) is the part of the prior knowledge
- Your response is
  - Then, $P(\theta|D)$ is the conclusion influenced by the data and the prior knowledge?
- Bayes says
  - Yes, and it will be our future prior knowledge!

$$\boldsymbol{P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}}$$

$$\boldsymbol{Posterior = \frac{Likelihood \times Prior\ Knowledge}{Normalizing\ Constant}}$$

More Formula from Bayes Viewpoint



Nice match to the range!

- $P(\theta|D) \propto P(D|\theta)P(\theta)$
  - $P(D|\theta) = \theta^{a_H}(1-\theta)^{a_T}$
  - $P(\theta) =????$
- We need to represent the prior knowledge well
  - So, the multiply goes smooth and does not complicate the formula
- Bayes says
  - Why not use the Beta distribution?
  - $P(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}, B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \Gamma(\alpha) = (\alpha-1)!$
- Your response is
  - Wow convenient!
  - $P(\theta|D) \propto P(D|\theta)P(\theta) \propto \theta^{a_H}(1-\theta)^{a_T}\theta^{\alpha-1}(1-\theta)^{\beta-1}$
    $= \theta^{a_H+\alpha-1}(1-\theta)^{a_T+\beta-1}$
  - Also, you notice one interesting face from the above formula…

Maximum a Posteriori Estimation

- Billionaire says
  - Hey! Stop! I am here!
  - So, you are talking about the formula
  - I want the most probable and more approximate θ
- Your response is
  - We are there.
  - Previously in MLE, we found θ from $\hat{\theta} = argmax_{\theta} P(D|\theta)$
    - $P(D|\theta) = \theta^{a_H}(1-\theta)^{a_T}$
    - $\hat{\theta} = \frac{a_H}{a_H + a_T}$
  - Now in MAP, we find θ from $\hat{\theta} = argmax_{\theta} P(\theta|D)$
    - $P(\theta|D) \propto \theta^{a_H+\alpha-1}(1-\theta)^{a_T+\beta-1}$
    - $\hat{\theta} = \frac{a_H+\alpha-1}{a_H+\alpha+a_T+\beta-2}$
  - The calculation is same because anyhow it is the maximization

Conclusion from Anecdote

- Billionaire says
  - Wait you and Bayes!
  - Who is right? The numbers are different!
- Bayes says
  - Not really… if you give us enough money to replicate the game!
- You say
  - Yes! If a_H and a_T become big, α and β becomes nothing…
- Billionaire says
  - Enough talking
  - Still, α and β are important if I don't give you more trials
  - Who decides α and β?
- Bayes and you say
  - Well… maybe grad students? =)

MLE

$$\hat{\theta} = \frac{a_H}{a_H + a_T}$$

MAP

$$\hat{\theta} = \frac{a_H + \alpha - 1}{a_H + \alpha + a_T + \beta - 2}$$