Bayesian Estimation vs Maximum Likelihood Estimation

Parameter Estimation: Observed data x depends on unknown parameters α through pdf

$$x \sim f(x;\alpha)$$

Brief MLE

Log-likelihood Estimation $\Rightarrow \mathcal{L}(\theta) = \log P(\text{Data}|\theta)$

So, θ is estimated by maximizing the log-likelihood

Yani $\Rightarrow$ $\theta_{MLE} = \arg\max_\theta \mathcal{L}(\theta)$

Bayesian Estimation :

Given model $f(x \mid \alpha)$ and prior $f(\alpha)$,
find posterior pdf $f(\alpha \mid x)$

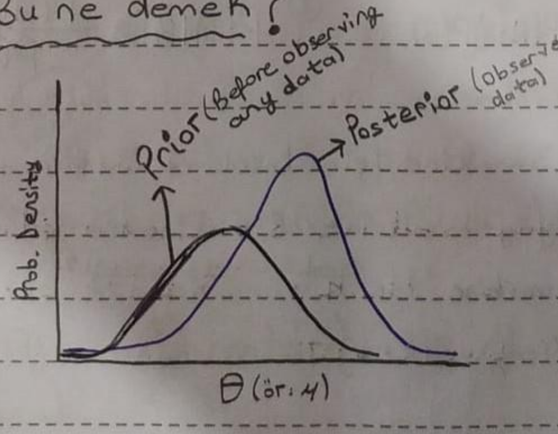
Baye's Rule : $f(\alpha \mid x) = f(x \mid \alpha) \cdot f(\alpha) \,/\, f(x)$

Ne demiştik? Frequentists use MLE to obtain a point estimation of the parameters θ.

Instead of Point Estimate, Bayesians estimate full posterior distribution of the parameters using Bayes formula.

$$P(\theta \mid Data) = \frac{P(Data \mid \theta)\, P(\theta)}{P(Data)}$$

Bu ne demek?



Yani; Bayesyen parametrelere ait Posterior dağılımı tahmin eder.
(Aslında parametreye ait dağılımı tahmin ediyoruz)

Peki birçok muhtemel θ değeri varken en iyisini nasıl bulacağız?

Bu durumda, aklımıza MLE'nin yaptığı gibi bir point estimate yapabilir miyiz diyoruz.

Tam bu noktada Maximum A Posteriori (MAP) estimation devreye girer.

↓

Tries to find Point estimates of θ that maximize the Posterior Distribution.

$$\theta_{MAP} = \arg\max_{\theta} P(\theta \mid Data)$$

Bayes Ağları (Bayesian Network)

Koşullu olasılık formülü: ① $P(A \mid B) = \dfrac{P(B \mid A) \cdot P(A)}{P(B)}$

$\star$ $P(A \cap B) = P(B \mid A) P(A) = P(A \mid B) P(B)$ ② $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$

ÖR: Yay üreten 3 makine için bozuk ve üretim oranları veriliyor. Eğer yay bozuk ise 3. makineden çıkma olasılığı nedir?

	Bozuk %	üretim oranı %
1. makine	2	35
2. "	1	25
3. "	3	40

$$P(B) = \frac{35}{100} \cdot \frac{2}{100} + \frac{25}{100} \cdot \frac{1}{100} + \frac{40}{100} \cdot \frac{3}{100} = \frac{215}{10000}$$

$$P(3.\text{makine} \mid B) = \frac{P(3.\text{mak}) \cdot P(B \mid 3.\text{mak})}{P(B)} = \frac{\frac{40}{100} \cdot \frac{3}{100}}{\frac{215}{10000}} = \frac{120}{215}$$

If X and Y are dependent events, then conditional probability is $\Rightarrow P(X|Y) = P(X \cap Y) / P(Y)$

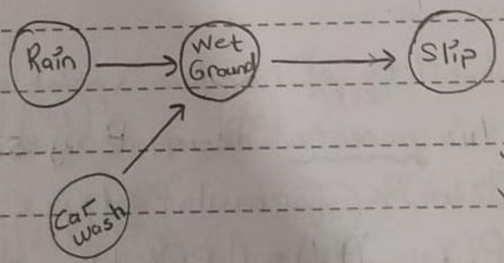
If are independent, then conditional prob. is $\Rightarrow P(X|Y) = P(X)$

★ Joint Probability $\Rightarrow P(A, B, C) \rightarrow$ The probability of event A, B, C occur at the same time (Kesişimleri yani)

★ $P(A \cap B) = P(A, B)$

ÖR. Bayesian Network:

Yolda yürürken kayma(slip) olasılığımız nedir?

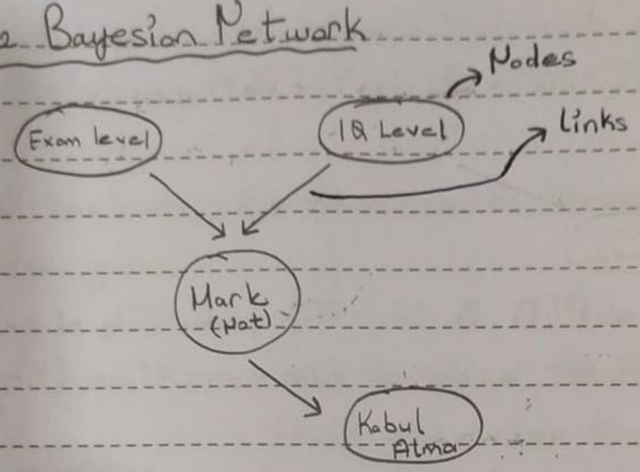


Joint probability of our 5 variables $\Rightarrow P(L, R, C, W, S)$

$$P(R, C, W, S) = P(R)P(C)P(W|C,R)P(S|W)$$

Bayes Ağları "Parents" mantığı ile çalışır. Arka sayfada açıklanacak.

Örn. Bayesian Network



Öğrencinin kabul alma olasılığına bakacağız.

Calculate the joint prob. of these 5 variables.

$P(a, m, i, e) = P(a|m) \cdot P(m|i,e) \cdot P(i) \cdot P(e)$

Görüldüğü üzere, Probability of random variable depends on his parents. Therefore, Bayesian Network formulated as;

* $P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i))$

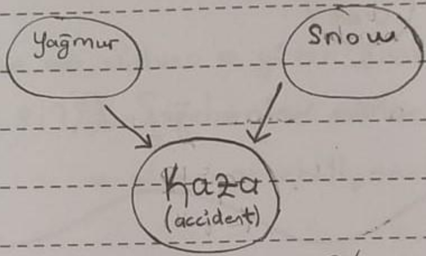
Yani; Örneğin;

- Öğrencinin "Mark"ı Exam level (Parent node) ve IQ Level (Parent node)'e bağlıdır.
- Öğrencinin kabul alması da Mark'a (Parent Node) bağlıdır.

Sonuç olarak;

Bir düğümün (Node) gerçekleşme olasılığı önceki düğümlerin ortak (joint) olasılığına bağlıdır.

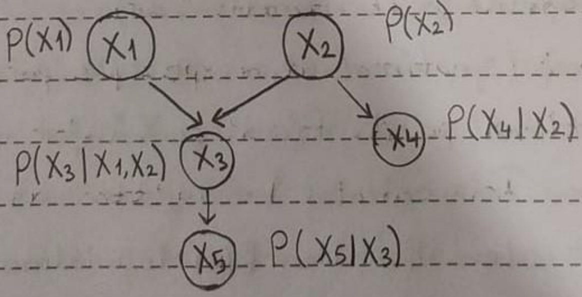
Bana şöyle bir örnek verseydi:



Kaza olasılığımız nedir? $P(y,s,k) = P(k \mid y,s) \cdot \underbrace{P(y) \cdot P(s)}_{\text{Bunların parent' yok.}}$

Ek: $P(A \mid B,C) = \dfrac{P(A \cap B \mid C)}{P(B \mid C)} = \dfrac{P(A \cap B \cap C)}{P(B \cap C)}$

Ek

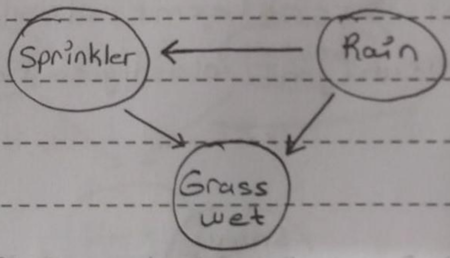


Ayrıca; $P(H \mid B) = \dfrac{P(B \mid H) \cdot P(H)}{P(B)}$ ⟹ P(B) olasılığı H olduğu ve olmadığı durumlardaki olasılıkların toplamına eşittir

$$P(H \mid B) = \frac{P(B \mid H) \cdot P(H)}{P(B \mid H) \cdot P(H) + P(B \mid \overline{H}) \cdot P(\overline{H})}$$

..../..../....

Bayes Ağı Son Örnek



Çimenlerin ıslak olma olasılığı? $P(r,s,g) = P(g|r,s) \cdot P(s|r) \cdot P(r)$