

WE ARE ALWAYS ON YOUR SIDE

KASAN
on your side



인공 지능, 기계 학습 세미나

Artificial Intellectual, Machine Learning

Chapter 1. 동기 및 기초

특허법인 가산
특허 6팀 김민석 변리사

● I. 서설

● II. 최우 추정법(Maximum Likelihood Estimation)

● III. 최대 사후 확률(Maximum Posteriori Estimation)

WE ARE ALWAYS ON YOUR SIDE

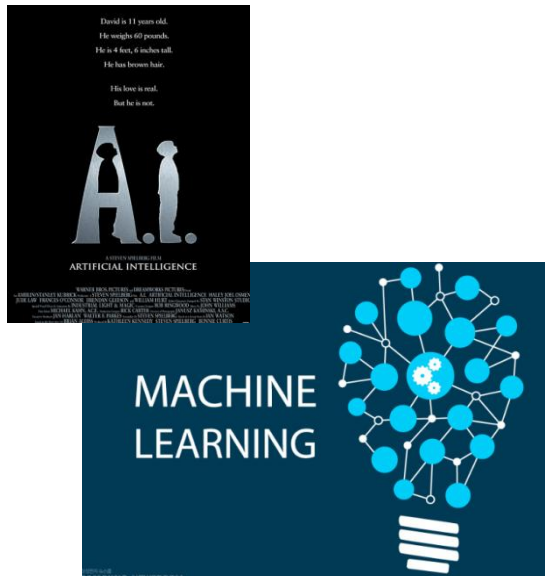
KASAN
on your side



I. 서설

✓ 인간의 지능으로 할 수 있는 사고, 학습, 자기계발 등을 컴퓨터가 할 수 있도록 하는 방법을 연구하는 컴퓨터 공학 및
정보기술의 한 분야로서, 컴퓨터가 인간의 지능적인 행동을 모방할 수 있도록 하는 것을 인공지능이라고 말하고 있다.

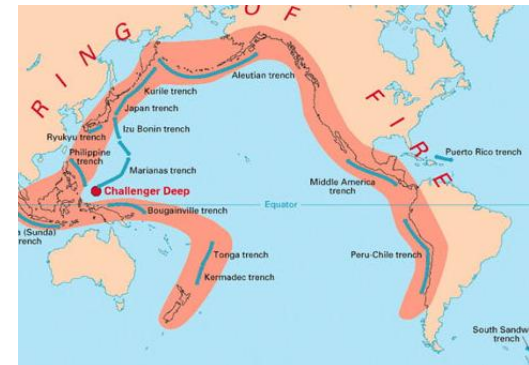
✓ 데이터의 생성 양·주기·형식 등이 방대한 빅 데이터들을 분석해 미래를 예측하는 기술을 일컫는다. 데이터를 수집, 분석해 미래를 예측한다는 점에서 빅데이터 분석과 유사하지만 컴퓨터 스스로가 방대한 데이터를 수집, 학습할 수 있다는 점에서 차이가 있다. 이는 인공지능(Artificial Intelligence)의 한 분야로, 빅 데이터 핵심 기술로 각광받고 있다.



I. 서설

◆ 자료의 범람

- ✓ 인터넷의 발달에 따른 데이터의 범람 : 관리 통제의 필요성
- ✓ 텍스트 데이터, 이미지 데이터, 시계열 데이터, 네트워크 데이터(SNS) 등



I. 서설

◆ 우리가 흔히 생각하는 인공지능



I. 서설

◆ 생각 보다 가까이 있는 머신 러닝의 예시

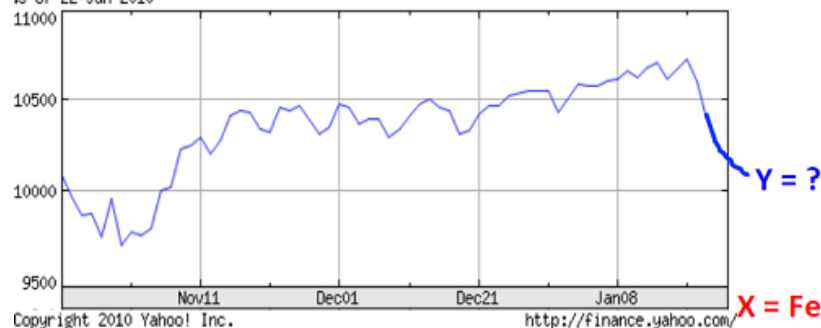
Document Classification



Sports
Science
News

Stock Market Prediction

INDU AVERAGE (DOO JONES & CO)
as of 22-Jan-2010



X = Feb01

Y = ?

Plate Num. Recognition



SNS Recommendation

Visit A Chartbanchachai and 7 others added you

Recommended Pages See All

글라라 Clara Lee
Hong Seok Yang and 2 other friends like her.
Like

SNL KOREA
Yong Hoon Choi and 3 other friends like this.
Like

TIME
Woojoo Lee and 4 other friends like this.
Like

People You May Know

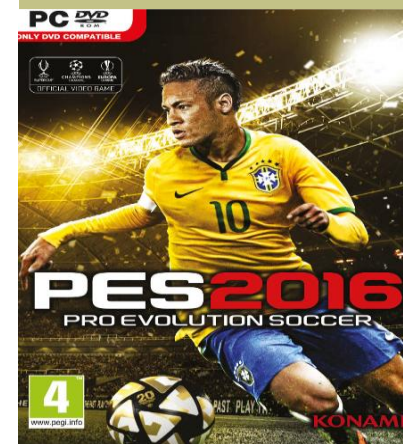
Alexander H. Levis, University Professor of Electrical, Connect
Doo-Hwan Bae, - Connect
Paul Davidsson, Professor at Malmö University Connect

See more »

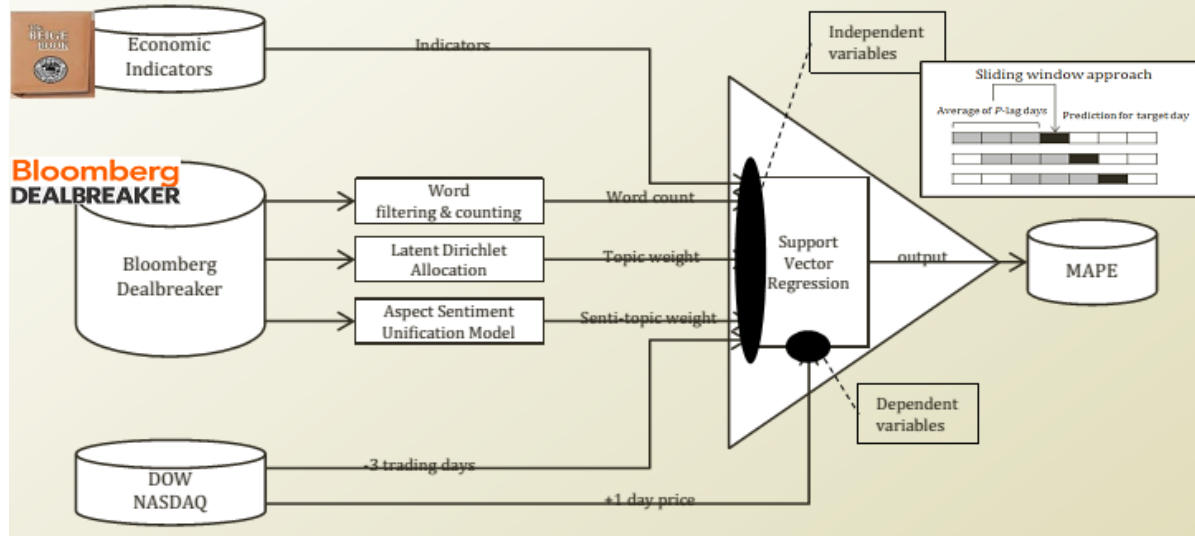
Ads You May Be Interested In

ENSEAD
The CCP Plenum Deciphered
Tsinghua-INSEAD EMBA Master Class on Mar. 7, Seoul.Discuss

Game

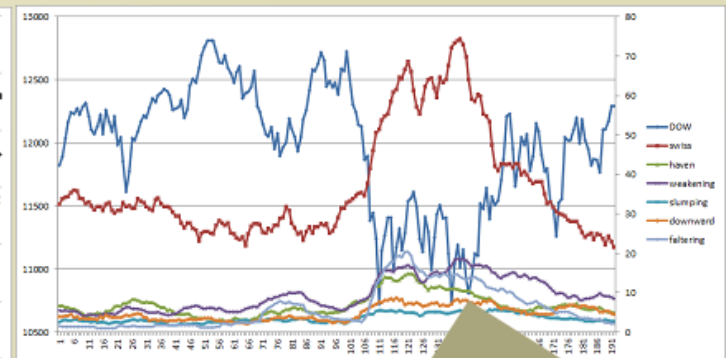
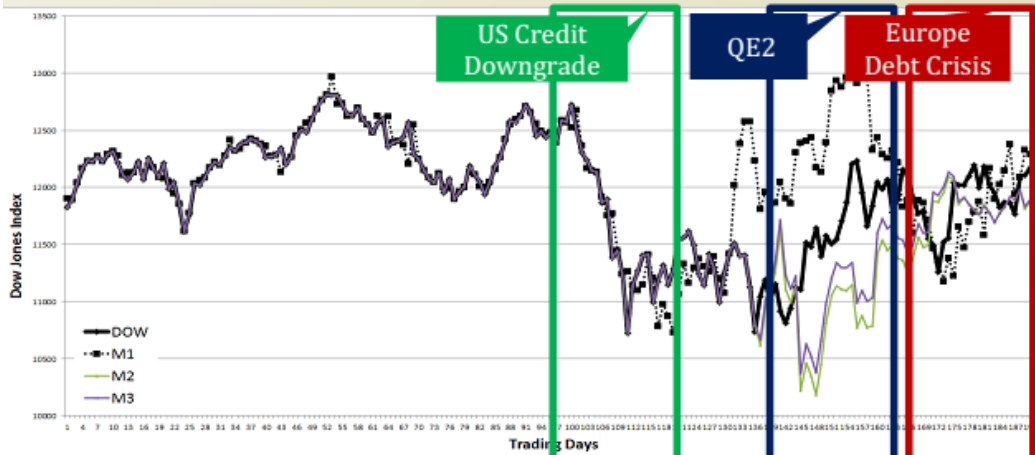


◆ Stock Market prediction



High Coefficients on Prediction

TopicWeight 26	TopicWeight 1	TopicWeight 7
-0.609	0.520	0.508
notes	obama	jun
moodys	republican	pence
swaps	republicans	na
treasuries	congress	swiss
versus	senate	chg
ratings	bill	francs
auCTION	barack	spa
default	lawmakers	fullyear
strategist	administration	nv
franc	democrats	dividend
twoyear	taxes	firstquarter
samp	white	ks
currencies	workers	paris
yen	democrat	reporting
swiss	obamas	tech



Heavy negative correlation between
"swiss" and DJIA

I. 서설

◆ AI in Foot ball Game

12. Lee Young-Pyo
수비수/필백 (왼쪽) / D/WB (R) - 원주버

능력치 프로필

신상 정보

종번호 12 나이 35
23.4.1977

소속팀 - 원주버

A매치 127경기 ... 5골
U-21 16경기 출장 2골

€2400 주급 €1800 세후 급여 €31.12.13 연로 €27.5만 몸값

66 kg 체중 176 cm 신장 양 발 발 쓰는 발

능력치

기술	정신	신체
개인기 17	공격 위치선정 10	균형 감각 8
골 결정력 3	대담성 10	몸싸움 7
드리블 16	리더십 9	민첩성 15
볼 트래킹 14	수비 위치선정 12	순간 속도 14
일대일 마크 11	승부욕 18	점프력 9
장거리 스토킨 9	예측력 14	주력 12
중거리 슛 3	적극성 10	지구력 13
코너킥 6	집중력 10	타고난 체력 19
크로스 10	창조성 9	골키퍼 등급 3
태클 12	전제성 8	
패스 12	침착성 16	
페널티킥 4	팀워크 14	
프리킥 4	판단력 13	
헤딩 9	활동량 14	

선수 데이터

Assistant Reports

Ask To Pick Show Filters Style Balanced Strategy Counter

Inf	Name	B Pos	Role	Duty	Ability
1	Daniel Vitali	GK	Goalkeeper	Defend	★★★★
2	Alexander Dominguez	GK	Goalkeeper	Defend	★★★★
3	Máximo Banguera	GK	Goalkeeper	Defend	★★★★
4	Niccolò Pissacco	D (R)	Full back	Support	★★★★
5	Pedro Velasco	D (R)	Full back	Support	★★★★
6	Luis Checa	D (L)	Full back	Support	★★★★
7	Diego Calderín	D (L)	Full back	Attack	★★★★
8	Jayro Campos	D (C)	Central defender	Covering	★★★★
9	Jorge Quaque	D (C)	Central defender	Stopper	★★★★
10	Eduardo Morante	D (C)	Central defender	Covering	★★★★
11	Frickson Erazo	D (C)	Central defender	Covering	★★★★
12	Segundo Castillo	DM	Anchor man	Defend	★★★★
13	Renato Ibañez	M (R)	Winger	Attack	★★★★
14	Walter Ayala	M (L)	Winger	Attack	★★★★
15	Jaime Zamora	M (C)	Central midfielder	Support	★★★★
16	Edison Mendieta	M (C)	Central midfielder	Support	★★★★
17	Patricio Urrutia	M (C)	Central midfielder	Support	★★★★
18	Diego Minda	M (C)	Central midfielder	Defend	★★★★
19	Christian Noboa	M (C)	Central midfielder	Support	★★★★

Field view showing player positions and stats on a soccer field.

팀 데이터

수원 FC vs 서울 FC

수원: 서울

2012. 4월 1. 일
날씨: 맑음 12C
심판: 강경환

수원 빅버드 스타디움

수원 FC lineup:

Goalkeeper: 1 (Lee Sa)

Defenders: 13 (Seo Ju), 8 (Lee Sa), 15 (Hong Suk), 9 (Oh Sang Eun)

Midfielders: 23 (Edo Bonar), 4 (Kwak Goo), 29 (Kwak Hee Ju), 14 (Oh Beom)

Attackers: 10 (Daniel), 11 (Molina), 15 (Chen Ryoo), 16 (Hu Goo Sung), 18 (Lee To), 12 (Park In)

서울 FC lineup:

Goalkeeper: 1 (Lee Sa)

Defenders: 11 (Molina), 15 (Chen Ryoo), 16 (Hu Goo Sung), 18 (Lee To), 12 (Park In)

Attackers: 10 (Daniel), 11 (Molina), 15 (Chen Ryoo), 16 (Hu Goo Sung), 18 (Lee To), 12 (Park In)

팀 vs 팀 데이터

경기 정보

리그 순위

현재 시즌

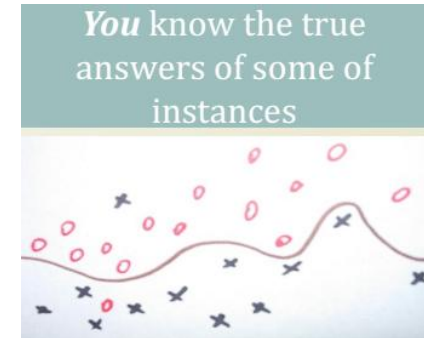
순위	팀명	경기	승	무	패	득점	실점	승점
1	전북	32	21	7	4	76	41	35
2	수원	32	20	7	5	84	49	35
3	서울	33	18	6	9	79	48	31
4	포항	33	18	5	10	55	44	11
5	경남	33	13	12	8	58	53	5
6	인천	33	11	6	16	55	54	1
7	제주	33	9	12	12	48	63	-15
8	상주	33	10	7	16	46	69	-23
9	부산	33	9	9	15	40	56	-16
10	대구	33	9	6	18	47	66	-19
11	전남	33	7	7	19	48	74	-26
12	성남	33	7	6	20	56	75	-19

리그 데이터

◆ 머신 러닝의 분류

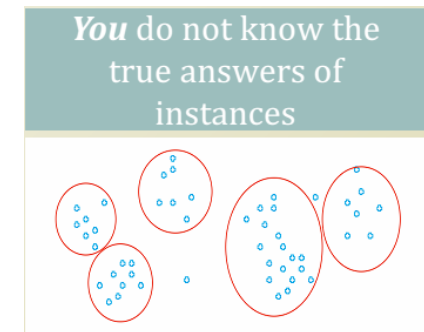
◆ Supervised Learning

- ✓ 각 객체(instance)의 결과(answer)를 설계자가 모두 알고 있으며, 해당 내용을 참조하여 기계 학습을 수행함.
- ✓ 주어진 데이터의 라벨을 매핑하는 함수



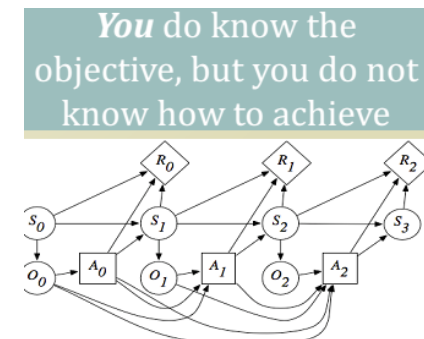
◆ Unsupervised Learning

- ✓ 각 객체(instance)의 결과(answer)를 설계자가 모두 알고 있지않아 기계가 직접 내용을 추론함
- ✓ 요약, 정리, 군집화 등



◆ Reinforcement Learning

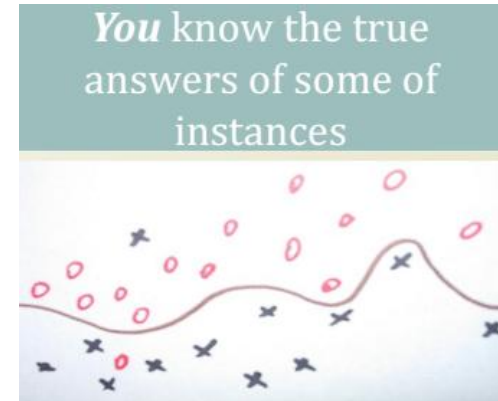
- ✓ 어떠한 행동을 취했을 때, 해당 행동에 대한 포상이 최대가 되는 방향으로 기계학습을 수행하는 것



◆ Supervised Learning

◆ 기계학습의 방법

- ✓ 각 객체(instance)의 결과(answer)를 설계자가 모두 알고 있으며, 해당 내용을 참조하여 기계 학습을 수행함.
- ✓ You know the true value, and you can provide examples of the true value



◆ 비교

- ✓ 스팸 필터링, 자동 등급 부여, 자동 분류
- ✓ Classification(분류)
 - ✓ Hit or miss, Ranking(A+, A0, B+, B0...), positive or negative
 - ✓ 연속적이지 않은 값에 대한 기계 학습
- ✓ Regression(회귀)
 - ✓ Value prediction, stock estimation
 - ✓ 연속적인 값에 대한 기계학습

◆ Unsupervised Learning

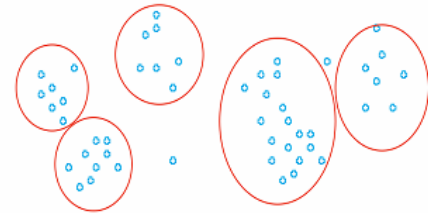
◆ 기계학습의 방법

- ✓ 각 객체(instance)의 결과(answer)를 설계자가 모두 알고 있지않아 기계가 직접 내용을 추론함
- ✓ You don't know the true value, and you cannot provide examples of the true value.

◆ 비교

- ✓ 설계자가 제시하는 기준(supervision) 없이 자료를 분석
- ✓ Discovering Cluster, latent factor(잠재 요소), graph structure
- ✓ Clustering
 - ✓ 방대한 신문 기사 중 주제를 요약
 - ✓ 군집화, 요약, 정리
- ✓ Filtering

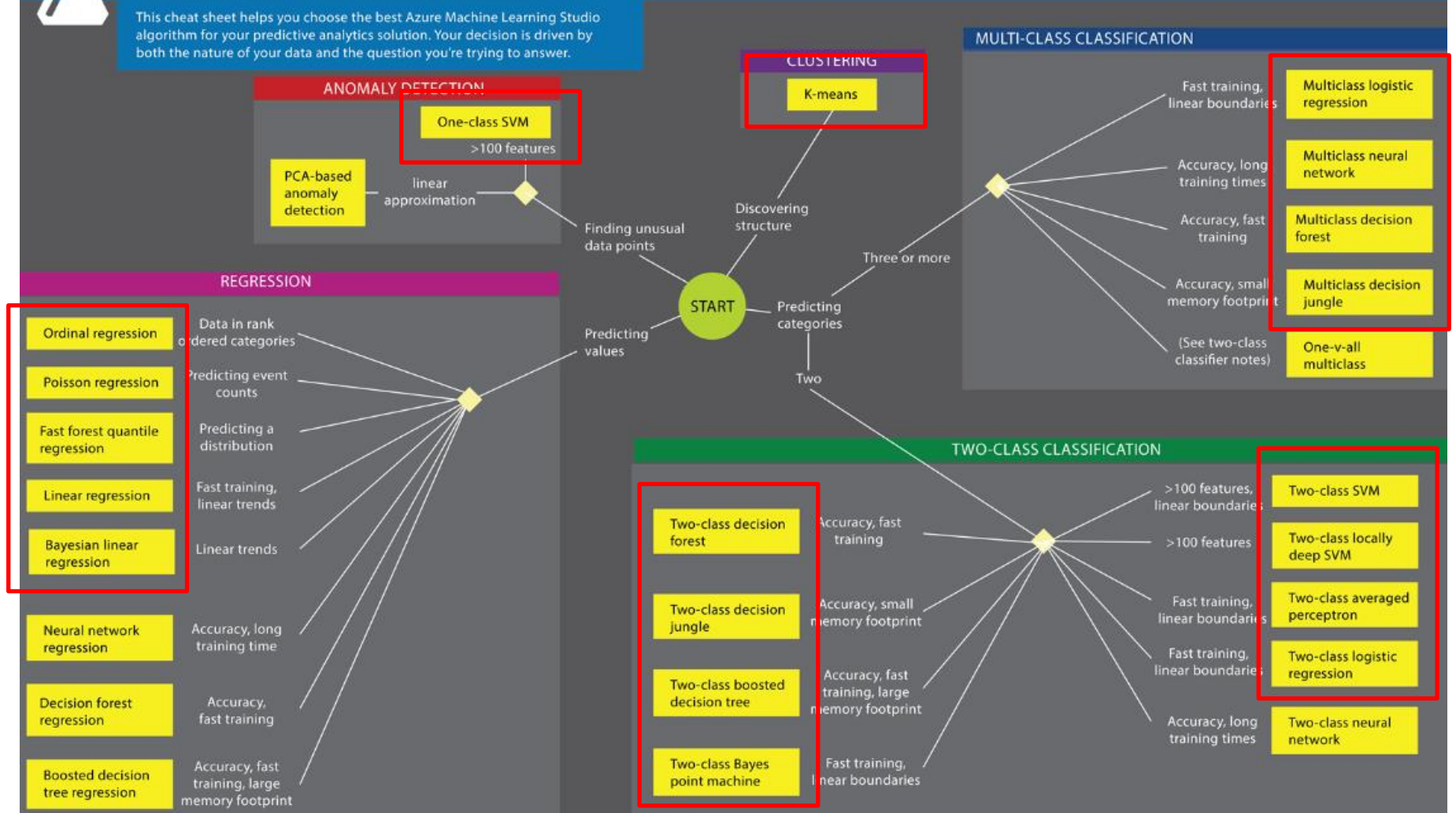
You do not know the true answers of instances

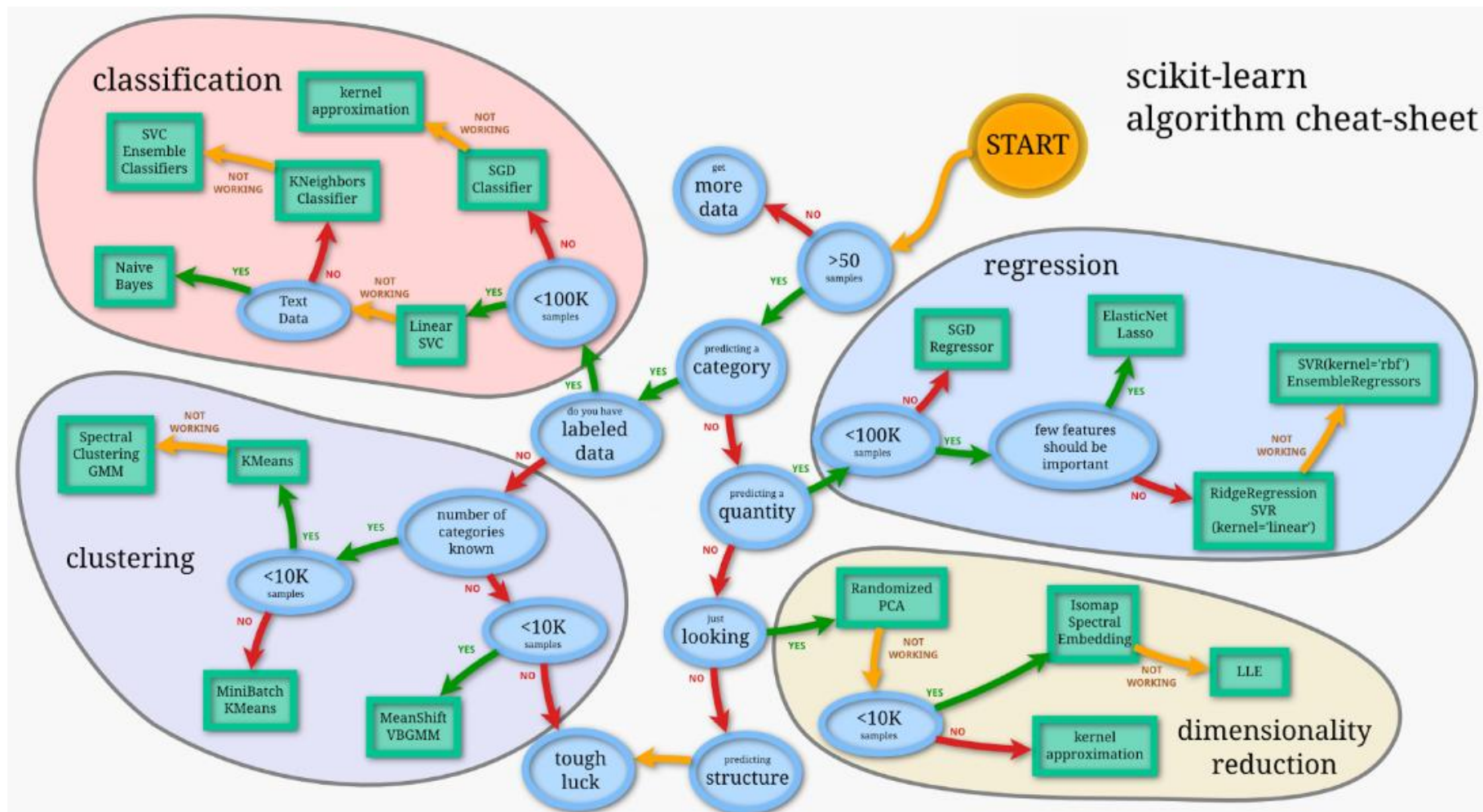




Microsoft Azure Machine Learning: Algorithm Cheat Sheet

This cheat sheet helps you choose the best Azure Machine Learning Studio algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the question you're trying to answer.







II. 최우 추정법(Maximum Likelihood method)

II. 최우 추정법

◆ 배경 지식 - 확률

- ✓ Philosophically, Either of the two
 - Objectivists assign numbers to describe states of events, i.e. counting
 - Subjectivists assign numbers by your own belief to events, i.e. betting
- ✓ Mathematically
 - A function with the below characteristics

$$P(E) \in R \quad P(E) \geq 0 \quad P(\Omega) = 1$$
$$P(E_1 \cup E_2 \cup \dots) = \sum_{i=1}^{\infty} P(E_i) \text{ when a sequence of mutually exclusive}$$

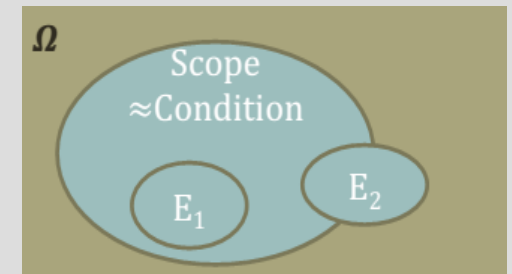
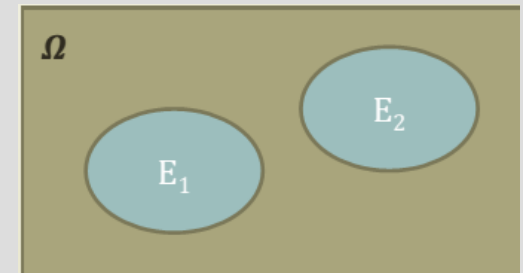
$$\text{if } A \subseteq B \text{ then } P(A) \leq P(B) \quad P(\emptyset) = 0 \quad 0 \leq P(E) \leq 1$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad P(E^C) = 1 - P(E)$$

- ✓ Conditional Probability

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

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

$$P(A) = \sum_n P(A|B_n)P(B_n)$$



II. 최우 추정법

◆ 배경 지식 - 확률 분포

- ✓ PDF(Probability Density Function)
- ✓ CDF(Cumulative Distribution)
- ✓ Normal Distribution

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Notation: $N(\mu, \sigma^2)$

Mean: μ

Variance: σ^2

- ✓ Beta Distribution

Supports a closed interval

- Continuous numerical value
- [0,1]
- Very nice characteristic
- Why?
 - Matches up the characteristics of probs

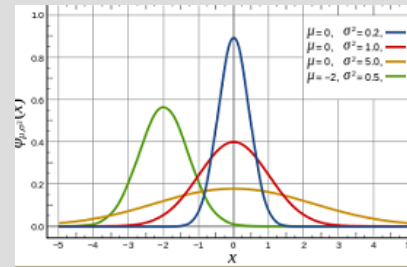
$$f(\theta; \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}, B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},$$

$$\Gamma(\alpha) = (\alpha-1)!, \alpha \in \mathbb{N}^+$$

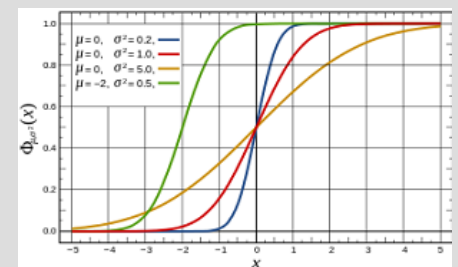
Notation: Beta(α, β)

Mean: $\frac{\alpha}{\alpha+\beta}$

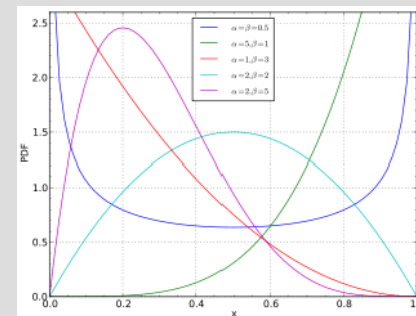
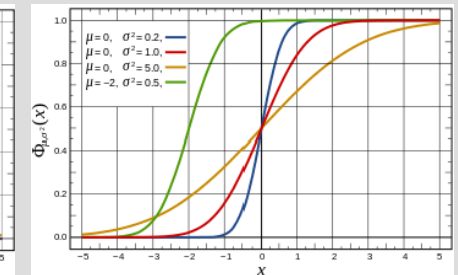
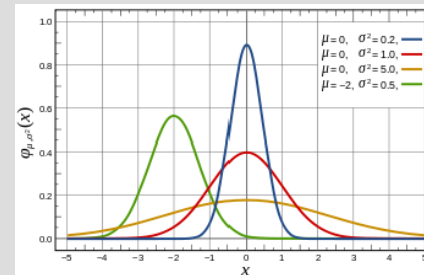
Variance: $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$



Probability Density Function
(PDF)= $f(x)$



Cumulative Distribution Function
(CDF) = $\int_{-\infty}^x f(x) dx$



II. 최우 추정법

◆ 압정 문제

- ✓ 임의의 압정을 던져서 압정이 꼬리(Tail)가 하늘로 향하는 경우 돈을 딸 수 있는 도박이 있다고 가정한다.
- ✓ 한 부자는 이러한 도박 경기에 임하기에 앞서 당신에게 조언을 구하고 있다.
 - ✓ 그는 꼬리가 하늘로 향하는 확률은 얼마인지에 대해서 당신에게 묻고 있다.
- ✓ 이때, 당신의 답변은?



Head



Tail

II. 최우 추정법

◆ 경험에 따른 판단

- ✓ 당신은 부자에게 압정을 몇회 던져 보라고 조언 할 수 있다.
- ✓ 부자는 총 5번의 압정을 던졌으며 이때
 - ✓ Head은 2번
 - ✓ Tail은 3번이 나왔다
- ✓ 당신이 부자에게 Tail이 나올 확률은 3/5라고 대답한다면 그 근거는 무엇인가?



Head



Head



Tail



Tail



Tail

II. 최우 추정법

◆ 배경 지식 - 확률 분포

- ✓ PDF(Probability Density Function)
- ✓ CDF(Cumulative Distribution)
- ✓ Binomial Distribution

$$f(\theta; n, p) = \binom{n}{k} p^k (1-p)^{n-k}, \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Notation: $B(n, p)$

Mean: np

Variance: $np(1-p)$

- ✓ Multinomial Distribution

- The generalization of the binomial distribution

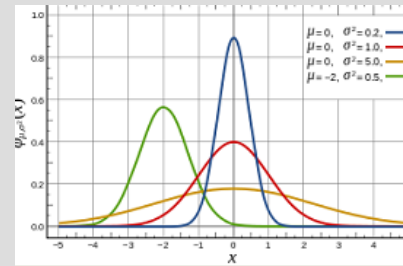
- Beyond yes/no
- Choose A, B, C, D, E, ..., Z
- Word selection, cluster selection,

$$f(x_1, \dots, x_k; n, p_1, \dots, p_k) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}$$

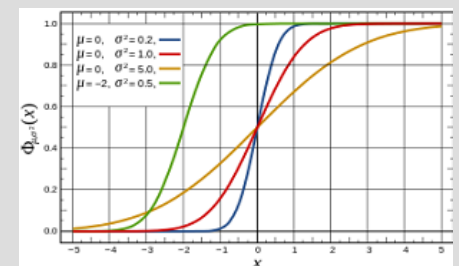
• Notation: $\text{Mult}(P), P = \langle p_1, \dots, p_k \rangle$

• Mean: $E(x_i) = np_i$

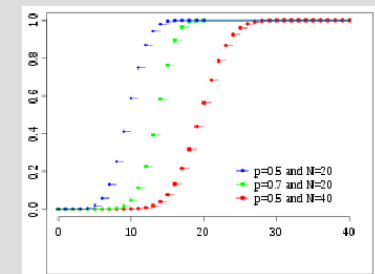
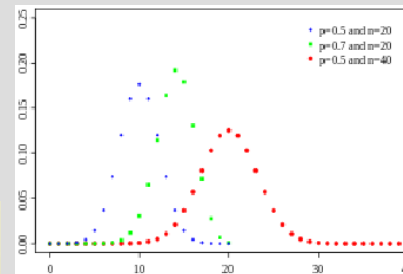
• Variance: $\text{Var}(x_i) = np_i(1-p_i)$



Probability Density Function
(PDF) = $f(x)$



Cumulative Distribution Function
(CDF) = $\int_{-\infty}^x f(x) dx$



II. 최우 추정법

◆ 이산 확률 분포 - Binomial Distribution

✓ 이산 확률 분포

- 복수의 독립적인 성공/실패 시험 중 일정한 횟수의 “성공”을 얻는 경우, 각각의 성공은 “ θ ”의 확률을 가짐
- Bernoulli experiment라고도 불림

✓ Independent Identically distributed(IID)

- 독립적인 이벤트 (Independent events) : 각각의 이벤트가 서로에게 영향을 주지 않으며 항상 정해진 이산확률 분포에 따라 결과를 얻음

✓ 압정 문제에 적용

$$P(H) = \theta, P(T) = 1 - \theta$$

$$P(HHTHT) = \theta\theta(1-\theta)\theta(1-\theta) = \theta^3(1-\theta)^2$$

Let's say

- D as Data = H,H,T,H,T

- $n=5$
- $k=a_H=3$
- $p=\theta$

- $P(D|\theta) = \theta^{a_H}(1-\theta)^{a_T}$

Makes order insensitive

n and p are given as parameters, and the value is calculated by varying k

$$f(k; n, p) = P(K = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\binom{n}{k} = \frac{n!}{k! (n-k)!}$$

◆ 최우 추정법 (Maximum Likelihood Estimation)

- ✓ 시퀀스 데이터 D는 aH의 Head 횟수와 aT의 Tail 횟수를 가진다고 할 때,
- ✓ 확률 분포를 구성하는 가장 타당한 확률 θ 를 구하는 것

One candidate is the **Maximum Likelihood Estimation (MLE)** of θ

- Choose θ that maximizes the probability of observed data

$$\hat{\theta} = \operatorname{argmax}_{\theta} P(D|\theta)$$

II. 최우 추정법

◆ 최우 추정법 (Maximum Likelihood Estimation) 값의 계산

✓ $\hat{\theta} = \operatorname{argmax}_{\theta} P(D|\theta) = \operatorname{argmax}_{\theta} \theta^{a_H} (1 - \theta)^{a_T}$

This is going nowhere, so you use a trick

- Using the log function

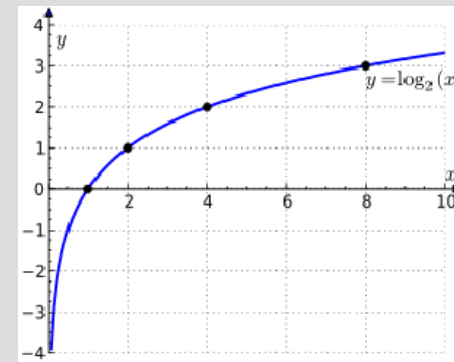
✓
$$\begin{aligned}\hat{\theta} &= \operatorname{argmax}_{\theta} \ln P(D|\theta) = \operatorname{argmax}_{\theta} \ln \{ \theta^{a_H} (1 - \theta)^{a_T} \} \\ &= \operatorname{argmax}_{\theta} \{ a_H \ln \theta + a_T \ln(1 - \theta) \}\end{aligned}$$

✓ Then, this is a maximization problem, so you use a derivative that is set to zero

- $\frac{d}{d\theta} (a_H \ln \theta + a_T \ln(1 - \theta)) = 0$
- $\frac{a_H}{\theta} - \frac{a_T}{1 - \theta} = 0$
- $\theta = \frac{a_H}{a_T + a_H}$
- When θ is $\frac{a_H}{a_T + a_H}$, the θ becomes the best candidate from the MLE perspective

✓ 결론 :

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



II. 최우 추정법

◆ 압정 문제

- ✓ 당신은 부자에게 MLE에 따라서 압정이 Tail이 나올 확률이 0.6이라고 보고 할 수 있다.
- ✓ 부자는 이를 수긍 하였으나 다음과 같이 묻는다
 - ✓ “자네가 계산을 하고 있는 동안 나에게 시간이 남길래 내가 압정을 몇 번 더 던져 보았네, 그랬더니 Head가 20번, Tail이 30번 나오더군. 이 결과가 우리의 전략을 바꾸는데 도움이 되는가?”
- ✓ 당신은 답하길
 - ✓ Tail이 나올 확률은 여전히 0.6입니다.
 - ✓ 부자 : 그렇다면 나는 그저 시간을 낭비한 것 뿐인가?



Head : 0.4



Tail : 0.6

5 times test vs 50 times test?

II. 최우 추정법

◆ Simple Error Bound - PAC learning

- ✓ 당신은 부자에게 반복된 시험은 추정의 에러를 줄일 수 있다고 말할 수 있다.

Right now, we have $\hat{\theta} = \frac{u_H}{a_H + a_T}, N = a_H + a_T$

Let's say θ^* is the true parameter of the thumbtack flipping for any error, $\varepsilon > 0$

We have a simple upper bound on the probability provided by Hoeffding's inequality

- ✓ 결론 :
$$P(|\hat{\theta} - \theta^*| \geq \varepsilon) \leq 2e^{-2N\varepsilon^2}$$

- ✓ 특정 에러 기대 값을 얻기 위한 N의 값을 구할 수 있다.

- ✓ Example To obtain $\varepsilon = 0.1$ with 0.01% case

- ✓ Probably Approximate Correct(PAC) learning

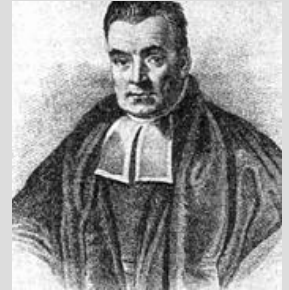


III. 최대 사후 확률(Maximum Posterior Estimation)

II. 최대 사후 확률

◆ 압정 문제

- ✓ 당신의 경쟁자(Bayes)가 부자에게 다음과 같이 말한다.
 - ✓ 경쟁자 : 잠시만요! Tail이 나올 확률이 0.6이 맞나요? 압정은 Head와 Tail이 존재하는데 그렇다면 확률은 50대 50이 되어야 하는 것이 아닌가요?
 - ✓ 부자 : 나도 처음에는 그렇게 생각하기는 했지. 그렇다면 실험 전 예측을 우리의 추정에 포함시킬 수 있는가?
- ✓ 데이터(D)가 주어지는 경우 “ θ ”가 사실일 확률



$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$
$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior Knowledge}}{\text{Normalizing Constant}}$$

$$P(D|\theta) = \theta^{a_H}(1 - \theta)^{a_T}$$

- ✓ 그렇다면 사전 정보(실험 전 예측)을 포함한 확률을 구하면?

III. 최대 사후 확률

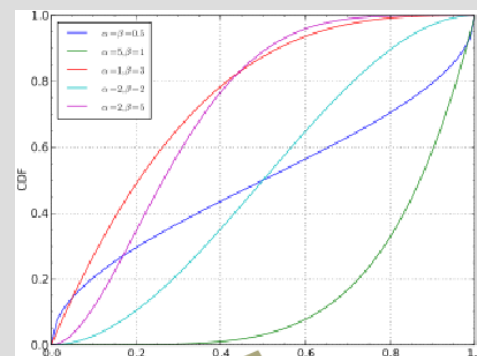
◆ Bayes의 이론에 따른 결론

- ✓ $P(\theta|D) \propto P(D|\theta)P(\theta)$
 - $P(D|\theta) = \theta^{a_H}(1 - \theta)^{a_T}$
 - $P(\theta) = \text{????}$

- ✓ 경쟁자(Bayes)는 말하기를

Why not use the Beta distribution?

$$P(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}, B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \Gamma(\alpha) = (\alpha-1)!$$



- ✓
$$\begin{aligned} P(\theta|D) &\propto P(D|\theta)P(\theta) \propto \theta^{a_H}(1 - \theta)^{a_T} \theta^{\alpha-1}(1 - \theta)^{\beta-1} \\ &= \theta^{a_H+\alpha-1}(1 - \theta)^{a_T+\beta-1} \end{aligned}$$

Also, you notice one interesting face from the above formula...

- ✓ $P(D|\theta) = \theta^{a_H}(1 - \theta)^{a_T}$ vs $P(\theta|D) \propto \theta^{a_H+\alpha-1}(1 - \theta)^{a_T+\beta-1}$

III. 최대 사후 확률

◆ 압정 문제

✓ Previously in MLE, we found θ from $\hat{\theta} = \operatorname{argmax}_{\theta} P(D|\theta)$

- $P(D|\theta) = \theta^{a_H}(1 - \theta)^{a_T}$
- $\hat{\theta} = \frac{a_H}{a_H + a_T}$

✓ Now in MAP, we find θ from $\hat{\theta} = \operatorname{argmax}_{\theta} P(\theta|D)$

- $P(\theta|D) \propto \theta^{a_H + \alpha - 1}(1 - \theta)^{a_T + \beta - 1}$
- $\hat{\theta} = \frac{a_H + \alpha - 1}{a_H + \alpha + a_T + \beta - 2}$

✓ 부자는 말하길

- ✓ MLE와 MAP의 결과가 다른데 그렇다면, 사전 정보를 추가하는 것 만으로도 확률이 변하는 것인가?
- ✓ 당산 : a_H 와 a_T 의 값이 아주 커지면 두 값은 결국 같아질 것 입니다.

MLE

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

MAP

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

WE ARE ALWAYS ON YOUR SIDE

KASAN
on your side



감사합니다

특허법인 가산
특허 6팀 김민석 변리사

◆ 참조문헌

- ✓ 카이스트 Open online course
 - <http://kaist.edwith.org/machinelearning/>
 - <https://www.youtube.com/channel/UC9caTTXVw19PtY07es58NDg/videos>(동영상 링크)
- ✓ 카이스트 SES LAB - 문일철 교수님
 - 인공지능, 기계학습 개론