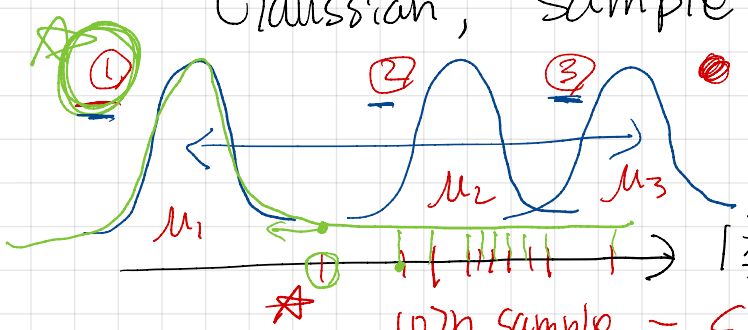
Lecture 14 | 중간고사 대비보강

* SVD의 보강설명. → 시험에 안냈음. 안심.
* MLE의 유도과정. → 시험에 안냈음. 안심.

SVD. 오늘 설명안함. → Principal Component Analysis
기회가 된다면 설명을 다음에. 주성분 분석. (PCA)

MLE: likelihood → 얼마나 일어날법하냐. 확률값.

Gaussian, sample → mean 값을 추정.



1축. 데이터.

$N(x; \mu, \sigma^2)$
모르는값.

10개 sample = {1.2, 1.8, 2.1, 1.7, 2.5, 3.2, 1.8, ⋯}
10개의 값.

⇒ Likelihood

$\mathcal{L}(\mu)$
~~$\mu \in \mathbb{R}$~~
$\mu \in \{\mu_1, \mu_2, \mu_3\}$
원소기호

sample이 일어날/관측될 확률.
= μ_1하에 sample이 관측될 확률

~~시험~~ $\boxed{\prod_{n=1}^{10} N(x_n; \mu_1, \sigma^2)}$ ← 각각의 sample이. 독립이니까

곱하겠다
$\sum$ 더하겠다.

$p(x_1, x_2, \cdots, x_n) = \prod_{k=1}^{n} p(x_k)$

$\prod_{k=1}^{3} k = 1 \times 2 \times 3$

$\sum_{k=1}^{3} k = 1+2+3$

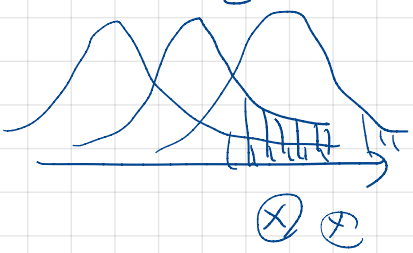
$$\mathcal{L}(\mu_1) = \prod_{n=1}^{10} N(x_n; \mu_1, \sigma^2)$$

$$\mathcal{L}(\mu_2) = \quad \mu_2$$

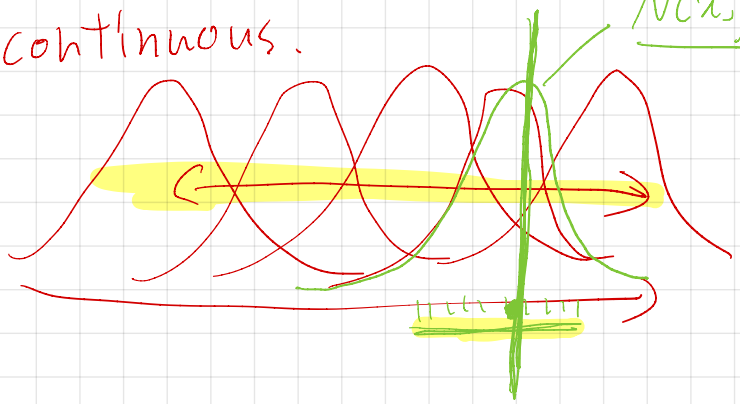
$$\mathcal{L}(\mu_3) = \quad \mu_3$$

$\mu \in \{\mu_1, \mu_2, \mu_3\}$

평균이 μ_1일때 sample이 관측될 확률



continuous.



$N(x; \hat{\mu}, \sigma^2)$ = 가우시안 가능도.

μ = 연속, 실수.

μ = ? MLE. $\mathcal{L}(\mu)$ max

$$\hat{\mu} = \frac{1}{N}\sum x_n$$ 평균

수학 개념 → 지반강의 유도과정. (시험 X)

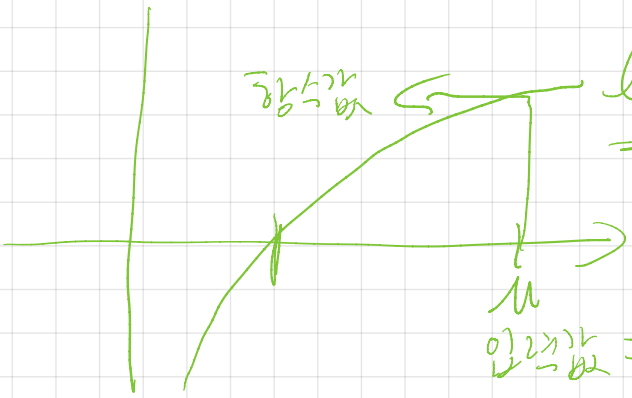
likelihood

정의. $$\mathcal{L}(\mu) = \prod_{n=1}^{N} N(x_n; \mu, \sigma^2)$$

log-likelihood를 씀.

$\log \mathcal{L}(\mu)$

$$\arg\max_{\mu} \mathcal{L}(\mu) = \arg\max_{\mu} \log \mathcal{L}(\mu)$$



log = 단조증가함수.

max value

arg max $\mathcal{L}$

arg max log $\mathcal{L}$

함수값

입력값 = 같은 위치에서 최대

$\mathcal{L}$, log $\mathcal{L}$

$\log \mathcal{L}(\mu)$ 이점 : ① 독립 = 곱 → 합

$\log x \cdot y = \log x + \log y$

② 확률분포가 많은경우 지수함수. exp

→ log를 통해 지수없앰.

~~공대~~, $\boxed{\text{기계공학과}}$ ← 실제 시스템

고등학교 $\log_{10} = \log$, $\log_e = \ln$

~~공대~~ 안령, $\log_{10}$ $\log$

$$\mathcal{L}(\mu) = \prod N(x_n; \mu, \sigma^2)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_1-\mu)^2}{2\sigma^2}\right)$$

$\exp(\cdot) = e^{(\cdot)}$ 복잡한거

log = 자연로그

$$\times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_2-\mu)^2}{2\sigma^2}\right)$$

e

$$\times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_3-\mu)^2}{2\sigma^2}\right)$$

$$\vdots$$

$$\log \mathcal{L}(\mu) = \boxed{\sum_{n=1}^{N}} \left(\boxed{\log \frac{1}{\sqrt{2\pi\sigma^2}}} - \frac{(x_n-\mu)^2}{2\sigma^2} \right)$$

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

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

$$= -\frac{1}{2} \log 2\pi\sigma^2 - \frac{(x-\mu)^2}{2\sigma^2}$$

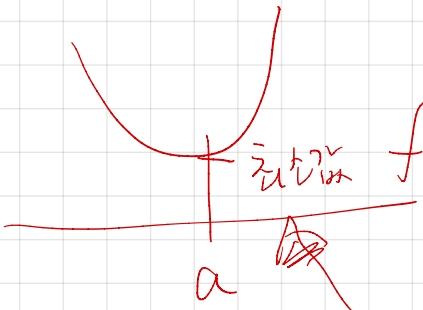
최대 μ

$$\boxed{\log \mathcal{L}(\mu)} = N \cdot \log \frac{1}{\sqrt{2\pi\sigma^2}} - \sum_{n=1}^{N} \frac{(x_n-\mu)^2}{2\sigma^2}$$

$$= -\frac{N}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2$$

min
max

미분 = 0 값 중에, min/max가 있을 수 있다.



최소값 $f(x) = (x-a)^2$

최소값 $\frac{d}{dx} f(x) = 0$

$2(x-a) = 0 \iff \boxed{x = a}$

최대

$$\frac{d}{d\mu} \log \mathcal{L}(\mu) = 0$$

상수

$$= \frac{d}{d\mu}\left\{ \cancel{-\frac{N}{2}\log 2\pi\sigma^2} - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 \right\}$$

$$= -\cancel{\frac{1}{\sigma^2}}\sum_{n=1}^{N}(x_n-\mu) \qquad \hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

$$= 0$$

$$\iff \sum_{n=1}^{N}(x_n-\mu) = 0 \iff \sum_{n=1}^{N} x_n = N\mu$$

SVM

Support Vector,

SVM = 학습데이터의 일부를 가지고 계산.

☆ 도움이 되겠죠?

$\boxed{\text{SVM}}$ → 「convex. optimization」 → 「Lagrangian dual problem」

↓

「해가 만족하는 조건 KKT condition」 → 「Support vector. (둘 중 하나는 0이 되어야 한다. complementary slackness)」

$a_n \times 0 = 0$: complementary slackness.

prediction $= \sum a_n \cancel{t_n K(x_n, x)} + \cdots$

곱해져 있어서 $a_n = 0$ 인

$a_1 = 0 \rightarrow x_1 = \cancel{SV}$

$a_n \neq 0 \rightarrow x_2 = SV$

---

primal problem: $\boxed{\max f_0(x)}$ s.t. $f_i(x) \geq 0$.

f_0, f_i convex

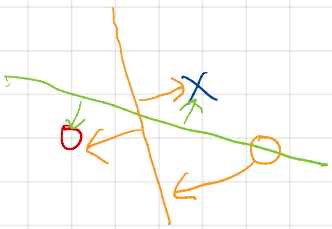
$\mathcal{L}$ dual problem: $\min_{\max} f_0(x) + \lambda_i f_i(x)$ s.t. $\lambda_i \geq 0$

$\mathcal{L}$ multiplier.

SVM.

primal problem: max (minimum 마진) s.t. (모든 샘플이 margin 밖에 위치하도록)

주어진 w, b로 결정된 boundary에 대해 가장 가까이 위치한 샘플까지의 거리



$$\boxed{\max} \left\{ \boxed{\min \left(\frac{1}{\|w\|} t_n (w^T \phi(x) + b) \right)} \right\}$$

s.t. $t_n(w^T\phi(x) + b) \geq 0$ 모든 sample이 자기영역에 들어갔다.

hard margin

linear model

$w^T\phi(x)+b$

$w \leftarrow \theta$

$\phi(x) \leftarrow x$

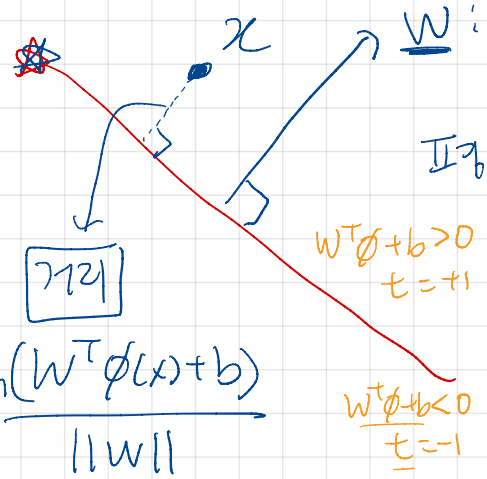
feature

b bias $\leftarrow \theta_0$

$\Leftarrow$ linear model

$\theta^T x + \theta_0$

$\boxed{w^T\phi(x)+b} \gtrless 0$



w : normal vector = 수직한.

평면을 정의

= 수직한 vector + bias 위치.

decision boundary

$t_n \in \{-1, 1\}$

$\boxed{\text{거리}}$

$$\frac{t_n(w^T\phi(x)+b)}{\|w\|}$$

$$\boxed{\begin{aligned} &\max \left\{ \min \frac{1}{\|w\|} t_n(w^T\phi(x)+b) \right\} \\ &\text{s.t.} \quad t_n(w^T\phi(x)+b) \geq 0. \end{aligned}}$$

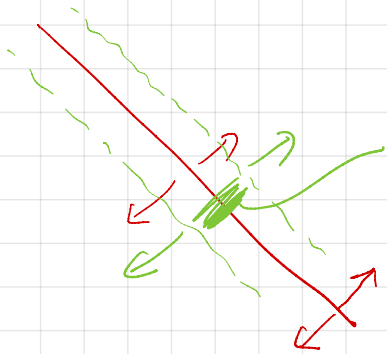
제설명이 부족

Bishop책.

+강하게.

margin보다도.

$$\text{s.t.} \quad t_n(w^T\phi(x)+b) \geq \min\{t_n(w^T\phi(x)+b)\} > 0$$



$$\min \frac{t_n(w^T\phi+b)}{\|w\|}$$

---

t@ 부교재.

등가 동일한문제인데 편의를위해 바꾸는 부분.

$$\text{margin} = \min \frac{t_n(w^T\phi(x)+b)}{\|w\|}$$

$\boxed{\text{normalize} \triangleq 1}$

$$\boxed{\max \frac{1}{\|w\|}}, \quad \text{s.t.} \quad \boxed{t_n(w^T\phi(x)+b) \geq 1}$$

$\rightarrow$ dual.

$\boxed{\min \frac{1}{2}\|w\|^2}$

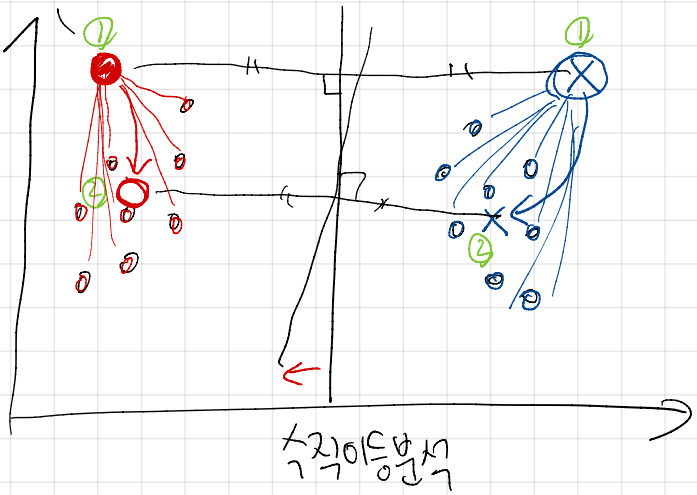
K-means

(lecture 7에 K-means Advanced에서 performance metric으로 나오는 inertia 안냈음.

$\boxed{\text{Inertia}} = \boxed{\text{mean}}$ squared distance

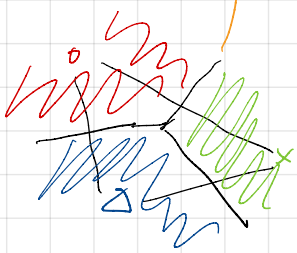
Supervision = 감독, 지시

ML
- supervised learning method ①
- unsupervised ... ② clustering



centroid, 평균

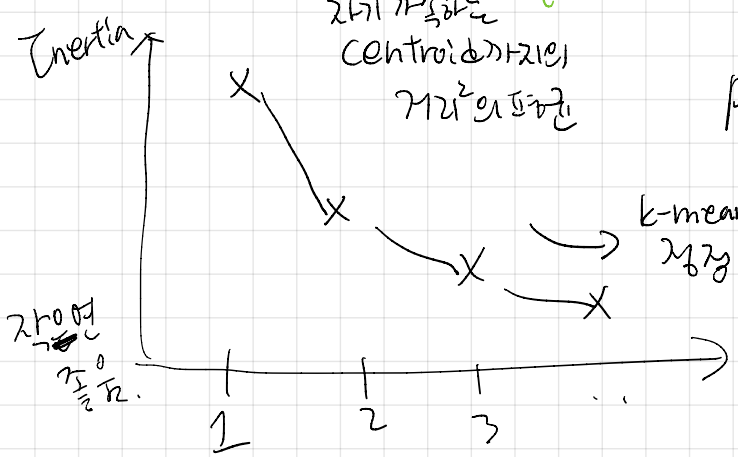
'voronoi tesselation'



16개의 거리

① Inertia = 평균 (빨간색이 제일가까운 데이터 (8개) 까지 빨간 centroid부터의거리 ^{2} 파란색 .. 제일가까운 데이터 (8개) 까지 파란 centroid부터의거리 ^{2})

② Inertia = 평균 (——— ^{2} — (~~8개~~) 까지 — cent. — 파란 ——— ^{2} 데이터 (~~8개~~) — — cent. —)



performance index. 지표.

k-means가 clustering을 잘하면, 잘나누면 점점 작은값 수렴.

SVM: Gaussian kernel 의 γ와 그 영향.

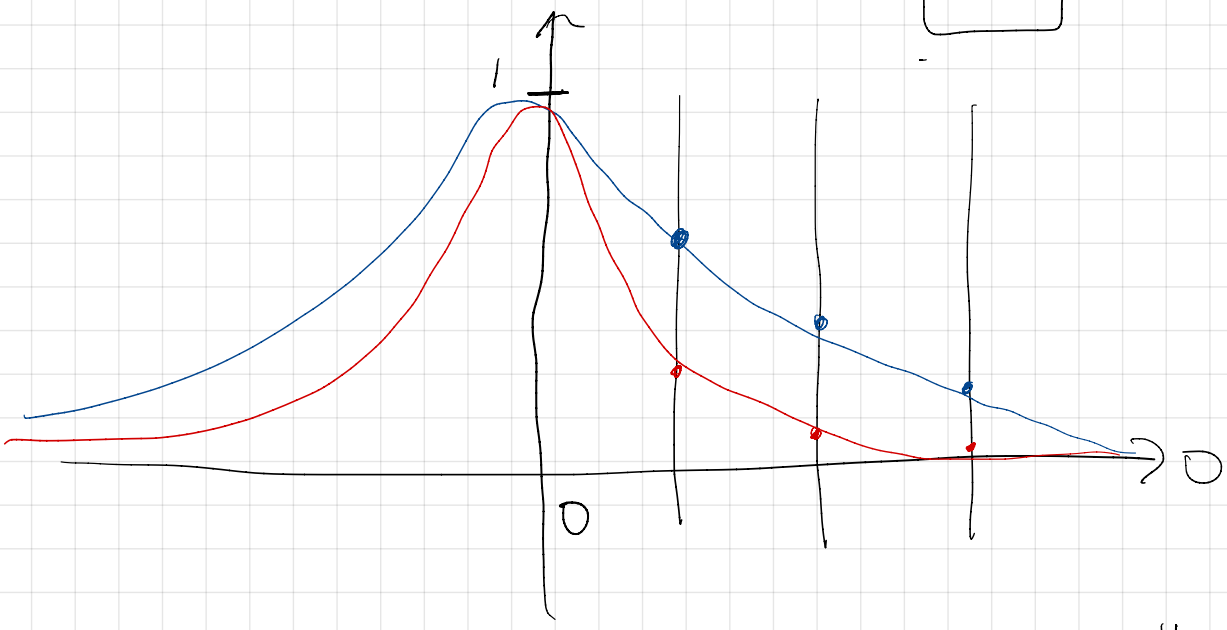
$$k(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2)$$

주어진 두 벡터에 대해 — 두 벡터 사이의 거리.

$$= \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right)$$

$$\exp(-x^2) = \frac{1}{e^{x^2}} \qquad \exp(-10x^2) = \frac{1}{e^{10x^2}}$$

x	1	2	3	4	분모	분모 역수	γ	σ^2
$\exp(-x^2)$	$\frac{1}{e}$	$\frac{1}{e^4}$	$\frac{1}{e^9}$	$\frac{1}{e^{16}}$	$\wedge$	$\vee$	작음 $\wedge$	더 큰값 더 좁
$\exp(-10x^2)$	$\frac{1}{e^{10}}$	$\frac{1}{e^{40}}$	$\frac{1}{e^{90}}$	$\frac{1}{e^{160}}$			큼	



$\|\mathbf{x} - \mathbf{x}'\|^2$