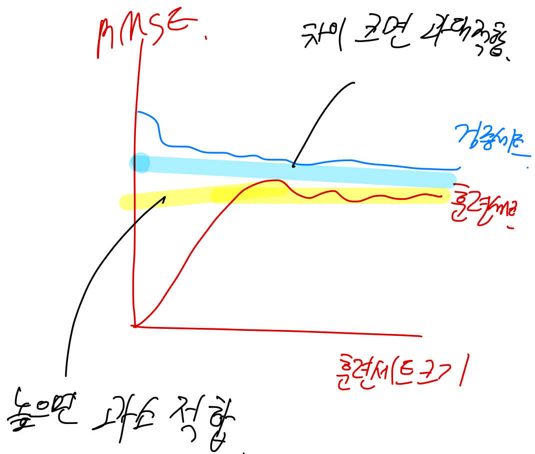
학습 곡선

비선형 데이터.

과대적합 & 과소 적합 확인하는 방법?

차수 (복잡도) 결정하는 방법?

⇒ 학습곡선



ideal

차수는 입력해야 함.

(but 경우의 수 별로 없을듯...)

특성 n

차원 d → $\frac{(n+d)!}{n!\,d!}$ 새로운 특성.

Vector norm $\|V\|_{p=1,2,\infty,\cdots}$

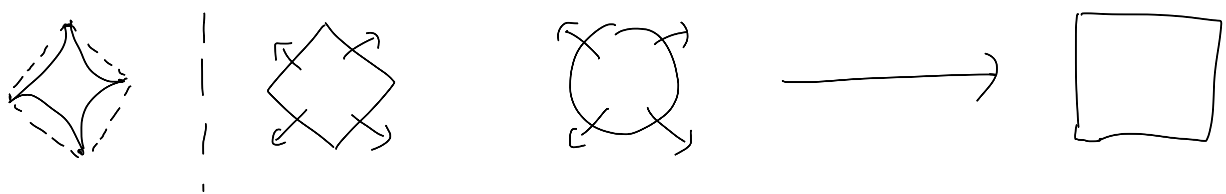
$p=1.$ $\|V\|_1 = |V_1| + |V_2| + \cdots |V_n|.$

$p=2.$ $\|V\|_2 = \sqrt{|V_1|^2 + |V_2|^2 + \cdots + |V_n|^2}.$

$p=\infty$ $\|V\|_\infty = \max |V_i|.$

$\|V\|_p = 1$ in $\mathbb{R}^2$

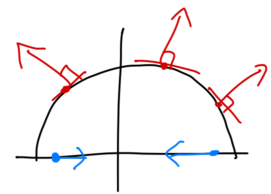
$p=\frac{1}{2}$ $p=1$ $p=2$ $p=\infty.$



"거리의 정의가 바뀌면 level set ($\|V\|=k$)의 모양이 바뀐다 !"

그래프를 바라보는 두 가지 관점.

i) Graph of function.
ii) Graph of level set.



$$\begin{cases} f : \text{function} \Rightarrow \nabla f : \text{the direction of which } f \text{ is increasing the fastest.} \\ F : \text{level set} \Rightarrow \nabla F : \text{normal vector of tangent plane.} \end{cases}$$

규제.

- 릿지 회귀. $J(\theta) = MSE + \underbrace{\frac{1}{2}\lambda \sum_{i=1}^{n} \theta_i^2}_{l_2\text{-norm}}$

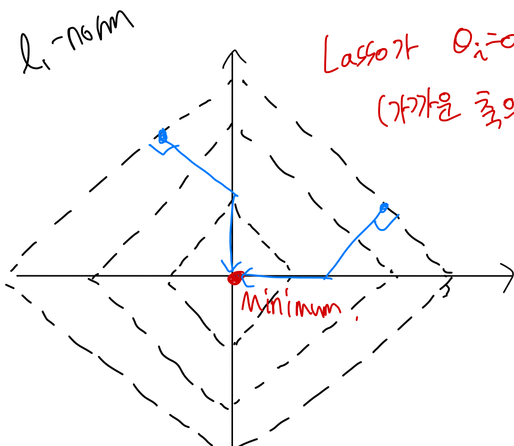
특징.

그래디언트 구하기 쉬움.

- 라쏘 회귀. $J(\theta) = MSE + \underbrace{\lambda \sum_{i=1}^{n} |\theta_i|}_{l_1\text{-norm}}$

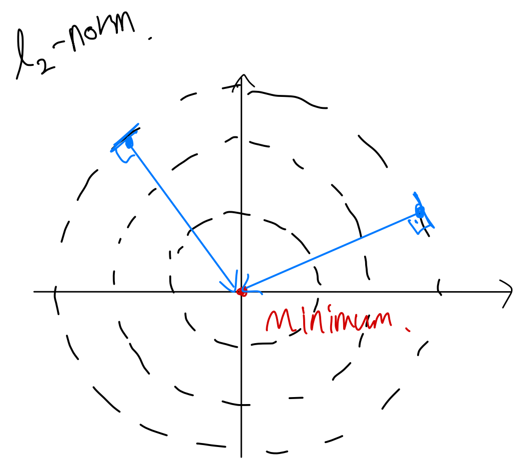
$\theta_i = 0$ (중요하지 않은 특성들)

걸러낼 수 있음.



Lasso가 $\theta_i = 0$ 많은 이유.

(가까운 축으로 먼저 떨어짐).



Note.

- 규제 전 데이터 전처리 해야함.
- 규제 항은 훈련하는 동안에만 사용.

$$\left(\begin{array}{l} \lambda \sum \theta_i^2 \\ \lambda \sum |\theta_i| \end{array} \right.$$

훈련에 사용하는 비용함수 & 테스트에 사용하는 비용함수. 다를 수 있음.

미분 가능해야함.

목적에 따라 다를 수 있음.

ex) 로지스틱 ⟶ 테스트에서는 정밀도/재현율로 평가.

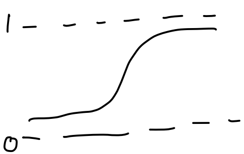
로지스틱 회귀

: 회귀 알고리즘을 분류에 사용.

$$X \longrightarrow \underline{\sigma}(\theta^T X) = \hat{p}. \qquad \left(\hat{p} : \begin{array}{l} 50\% \uparrow \longrightarrow O. \\ 50\% \downarrow \longrightarrow X. \end{array} \right)$$

sigmoid function $\sigma(t) = \dfrac{1}{1+e^{-t}}$

sigmoid



확률적인 해석 가능.

but $\underline{\sum_{i=1}^{n} \left(y - \sigma(\theta^T X)\right)^2}$

MSE를 Loss function으로 쓰기에 계산량 너무 많음.

⇒ 로지스틱에 맞는 Loss Function "Cross-entropy" 정의.

$X \longrightarrow \sigma(\theta^T X)$.

$X_1 = 5 \longrightarrow$ $sig(x_1, \theta)$ 확률로 5 입니다. $P(x_1) = sig(x_1, \theta) \uparrow$.

$x_2 = 4 \longrightarrow$ $sig(x_2, \theta)$ " . $P(x_2) = 1 - sig(x_2, \theta) \uparrow$.

$sig(x_2, \theta) \downarrow \Leftrightarrow 1 - sig(x_2, \theta) \uparrow$.

$\vdots$

$x_k \longrightarrow$. $P(x_k) = \begin{cases} sig(x_k, \theta), & y_k = 1 \\ 1 - sig(x_k, \theta), & y_k = 0 \end{cases}$.

5일 확률을 예측하는 회귀 모델.

$x_1, x_2, \cdots, x_n$: 독립시행 $\Rightarrow$ $\boxed{\text{We want to maximize } \prod_{k=1}^{n} P(x_k).}$

$$\text{maximize} \prod_{k=1}^{n} P(x_k)$$

$$\Longleftrightarrow \text{minimize} -\prod_{k=1}^{n} P(x_k).$$

$$\Longleftrightarrow \text{minimize} \log\left(-\prod_{k=1}^{n} P(x_k)\right).$$

log 함수의 성질.

- Monotone increase. ($x_1 < x_2 \longrightarrow \log x_1 < \log x_2$).
- $\log(1) = 0$. (확률이 1이면 오차 0).
- $\log(x_1 x_2) = \log x_1 + \log x_2$. (급격한 값의 변화를 완화시켜 줌).

  ($\times \longrightarrow +$).

$$\Rightarrow Loss = -\log\left(\prod_{k=1}^{n} P(x_k)\right)$$

$$= -\sum_{k=1}^{n} \log P(x_k)$$

$$\left(MSE : \sum_{k=1}^{n} \left(y_k - \sigma(\theta^T x_k)\right)^2\right)$$

in book.

$$J(\theta) = -\frac{1}{m}\sum_{k=1}^{n}\left[y^{(i)} \log(\hat{p}^{(i)}) + (1-y^{(i)}) \log(1-\hat{p}^{(i)}) \right]$$

$\Rightarrow$ l_2-norm을 이용한 Loss Function이 아니기 때문에 정규방정식 쓸 수 없음.

but gradient 생각보다 간단.

$$\left(\begin{array}{l} \sigma'(t) = \sigma(t)(1-\sigma(t)) \\ e^x \text{ 의미?} \end{array} \right)$$

- 로지스틱 회귀도 규제 가능.
- 결정 경계는 선형. ($\theta^T X = 0$).

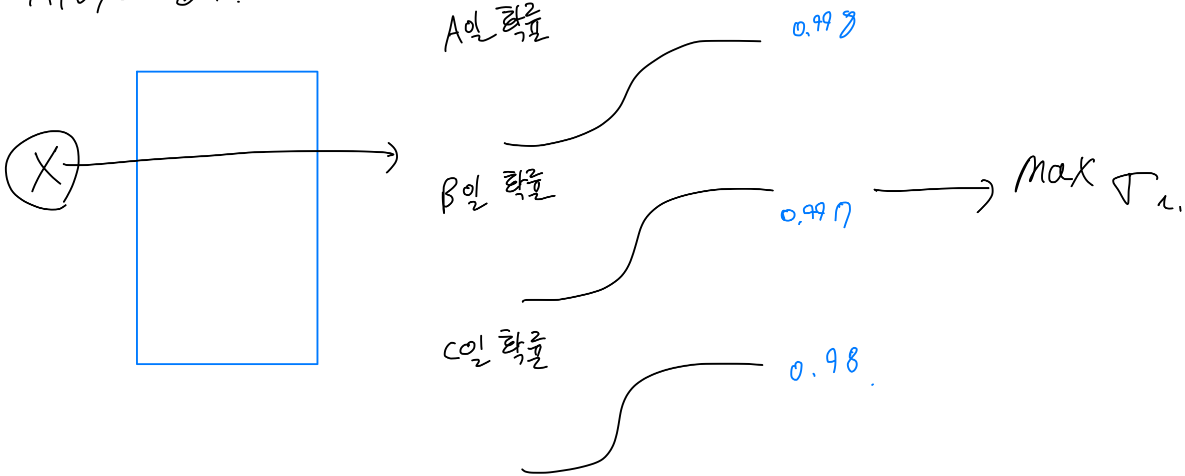
비선형 데이터의 경우,

KNN.
SVM
NN

소프트맥스 회귀.

: 다항 분류.

ex) A,B,C 분류.



다항 분류에서는 sigmoid 사용X.

A일 확률 $(\theta^{(1)})^T X$. 20

B일 확률 $(\theta^{(2)})^T X$. 70

C일 확률 $(\theta^{(3)})^T X$. 10.

$\Rightarrow$ $\dfrac{70}{20+70+10}$ 확률로 B,

차이를 더 부각시키는 방법?

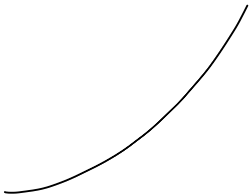
$$\boxed{\theta^T x \rightarrow e^{\theta^T x}}$$

$|5-3| = 2$

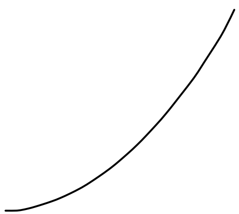
$\downarrow$

$|2^5 - 2^3| = 24$.

$$e^{(\theta^{(1)})^T X} = e^{s_1} \longrightarrow \overset{\text{A일 확률}}{\frac{e^{s_1}}{e^{s_1}+e^{s_2}+e^{s_3}}}$$



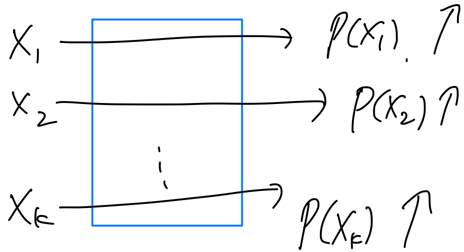
$$e^{(\theta^{(2)})^T X} = e^{s_2} \longrightarrow \overset{\text{B일 확률}}{\frac{e^{s_2}}{e^{s_1}+e^{s_2}+e^{s_3}}} \longrightarrow \max P_i$$



$$e^{(\theta^{(3)})^T X} = e^{s_3} \longrightarrow \overset{\text{C일 확률}}{\frac{e^{s_3}}{e^{s_1}+e^{s_2}+e^{s_3}}}$$

$$\boxed{\hat{P}_k = \frac{e^{s_k(x)}}{\sum_{j=1}^{k} e^{s_j(x)}}}$$

Loss function?



$\Rightarrow$ Cross-entropy 사용.

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left(\sum_{k=1}^{K} y_k^{(i)} \log\left(\hat{p}_k^{(i)}\right)\right).$$

y가 k_0 클래스에 속한다면

$$\cancel{y_1^{(i)} \log\left(\hat{p}_1^{(i)}\right)} + \cancel{y_2^{(i)} \log\left(\hat{p}_2^{(i)}\right)} + \cancel{\cdots} + \boxed{y_{k_0}^{(i)} \log\left(\hat{p}_{k_0}^{(i)}\right)} + \cancel{\cdots} + \cancel{y_K^{(i)} \log\left(\hat{p}_K^{(i)}\right)}.$$

$$\Rightarrow -\frac{1}{m}\sum_{i=1}^{m} \log\left(\hat{p}_{k_0}^{(i)}\right).$$

Cross-entropy 의 결정 경계?

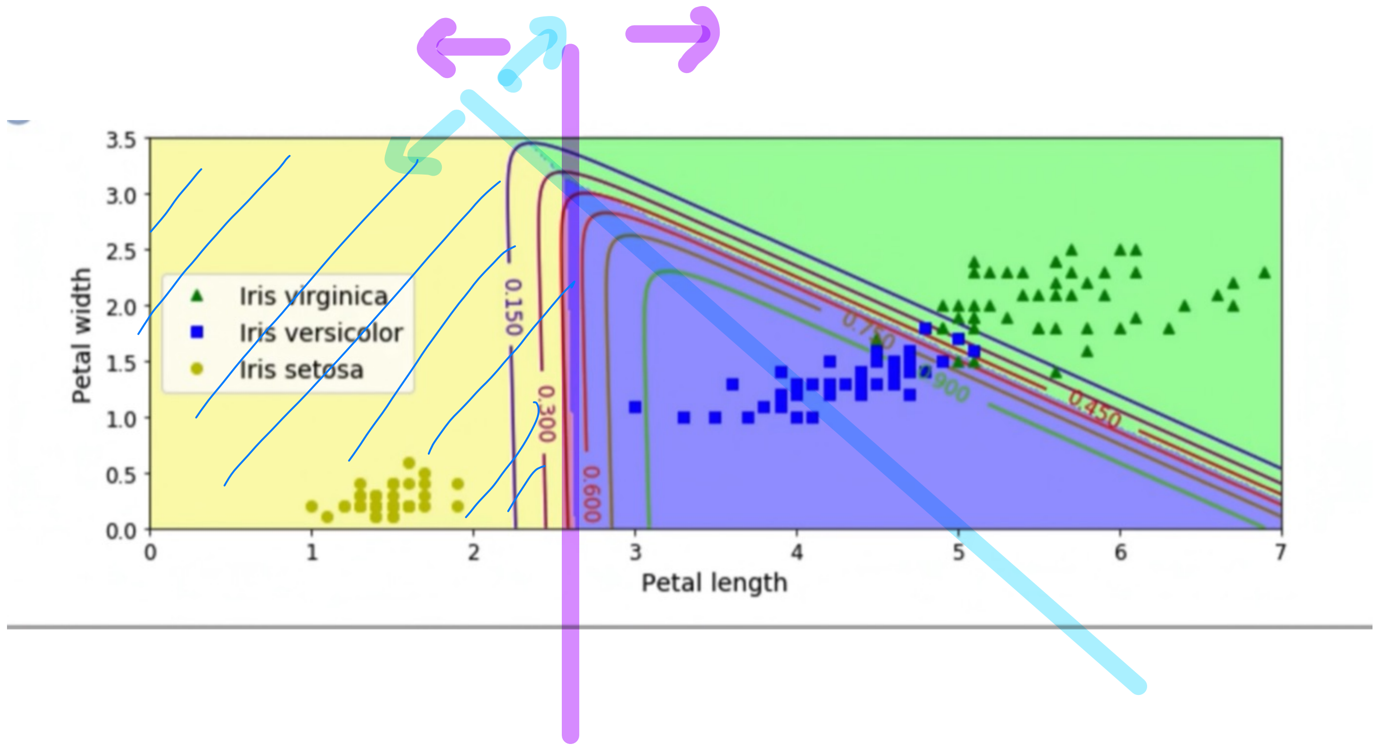
$$\frac{e^{s_1(x)}}{\sum_{j=1}^{k} e^{s_j(x)}} = \frac{e^{s_2(x)}}{\sum_{i=1}^{k} e^{s_j(x)}} \implies e^{s_1(x)} = e^{s_2(x)}$$

$$\implies s_1(x) = s_2(x)$$

$$\implies \theta_0^{(1)} + \theta_1^{(1)} x_1 + \cdots + \theta_n^{(1)} x_n = \theta_0^{(2)} + \theta_1^{(2)} x_1 + \cdots + \theta_n^{(2)} x_n.$$

$$\implies \boxed{(\theta_0^{(1)} - \theta_0^{(2)}) + (\theta_1^{(1)} - \theta_1^{(2)}) x_1 + \cdots + (\theta_n^{(1)} - \theta_n^{(2)}) x_n = 0.}$$

Softmax 의 결정경계는 선형!



Petal width
Petal length
Iris virginica
Iris versicolor
Iris setosa
0.150
0.300
0.600
0.750
0.900
0.450