$$\hat{y}_i = \underbrace{w_0}_{\text{offset}} + w_1 x_{i1} + \cdots + w_p x_{ip} = Xw = \langle w, x_i \rangle = \langle x_i, w \rangle$$

벡터의 내적 : $a \cdot b = a^T b = \sum_{k=1..d} a_k b_k$

x_{ij} = j^{th} feature of i^{th} Sample.

$$\text{Let } X = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}$$

$\hat{y} = Xw$ implies → 선형

$\hat{y}_i = w_1 + w_2 x_i + w_3 x_i^2 + w_4 x_i^3$ → 비선형

↳ Cubic Polynomial

residual $r_i = y - \hat{y}_i$, $r_i(w) = y_i - \hat{y}_i(w)$

벡터의 P차놈.

$$\|x\|_p = \left(\sum_{i=1..d} |x_i|^p \right)^{\frac{1}{p}}$$

Least Squares Estimation

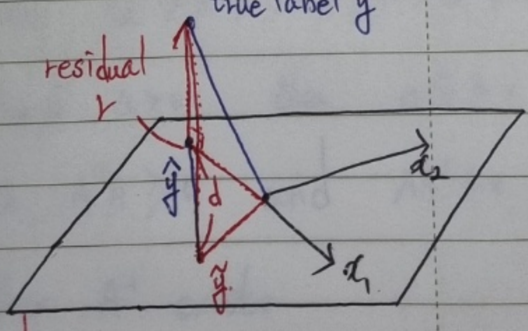
Choose $\underline{w}$ to minimize $\sum_{i=1}^{n} r_i(w)^2 = \|r\|_2^2$

$\|r\|_2 = \left(\sum_{i=1}^{n} r_i^2 \right)^{\frac{1}{2}}$ = l_2 (Euclidean Norm)

Least Square Estimation

1. Negative & Positive residuals are treated the Same
2. Mathmatically Convenient
3. Magnify large errors.
4. Nice geometric interpretation.
5. Consistent with Gaussian noise model $\underline{y} = Xw + \varepsilon$

Span : 구성벡터들로 형성할 수 있는 공간



↳ Span(col(X))

$\hat{y}$: x_1, x_2 가 span한 공간에서 y와 가장 가까운 점.

$\|\tilde{r}\|^2 = \|r\|^2 + \|d\|^2$

$$\text{Let } \hat{\underline{w}} = \underset{\underline{w}}{\operatorname{argmin}} \|r(\underline{w})\|^2$$

argument $\underline{w}$ that minimizes.

$\hat{r} = y - X\underline{\hat{w}} \Rightarrow X^T\hat{r} = X^T(y - X\underline{\hat{w}}) = 0$

↳ 그림에서 X와 r은 orthogonal

∴ $X^Ty = X^TX\underline{\hat{w}}$

선형독립 : 벡터 $V_1, V_2, V_3 \cdots V_n$이 있을 때, 모든 계수가 0인 경우를 제외하고 어떤 선형 조합으로도 0을 만들 수 없다면 독립.

Rank : 행렬의 열들 중 선형독립인 열들의 최대 개수 (행과 열의 Rank는 항상 동일)

Full Rank : 한 열이나 행에서 전부 다 선형독립인 벡터기저들을 가진 경우

if $rank(x) = min(n,p)$

Matrix only has an inverse if it is full rank

$X \in R^{n\times p}$ Assume that $n \geq p$, $rank(X) = p$

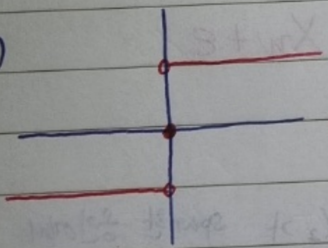
$\boxed{\text{↳ } X = \text{full rank}, \ X \text{ has an inverse}}$

$\boxed{rank(A) = rank(A^TA) = rank(AA^T)}$

∴ X^TX has an inverse

∴ $\underline{\hat{w}} = (X^TX)^{-1}X^Ty$

ex)



Sign function

$\tilde{y}_i = sign(\hat{y}_i) \in (-1,1)$

$$\hat{\underline{w}} = \underset{\underline{w}}{\arg\min}\ (\underline{y} - X\underline{w})^T(\underline{y} - X\underline{w})$$

$$= \underset{\underline{w}}{\arg\min}\ y^Ty - (X\underline{w})^T\underline{y} - y^TX\underline{w} + (X\underline{w})^T(X\underline{w})$$

$$= \underset{\underline{w}}{\arg\min}\ y^Ty - 2w^TX^Ty + w^TX^TX\underline{w} \qquad (w^TX^Ty = (y^TX\underline{w})^T)$$

양의 정부호 행렬 : 0이 아닌 모든 벡터 X에 대해 $X^TAX > 0$ $\quad A \succ 0$

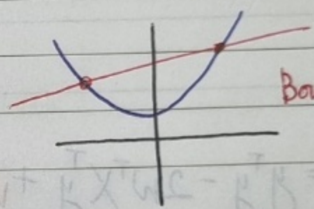
for $X \in R^{n\times p}$ with $n \geq p$, $\text{rank}(X) = p$

$(X^TX)^{-1}$ exists $=$ X^TX is positive definite (p.d.)

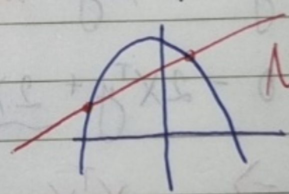
Convexity

함수 위의 두 점을 연결하는 선을 그었을 때 함수그래프 위만 지나가면 Convex

↳ For any point on the Curve, Can Compute its tangent



Bowl shape



Non-Convex

Properties of Positive Definite Matrices

- if $P \succ 0$ and $Q \succ 0$, then $P + Q \succ 0$
- if $Q \succ 0$ and $a > 0$ then $aQ \succ 0$
- For any A, $A^TA \succeq 0$ and $AA^T \succeq 0$
- if $A \succ 0$, then A^{-1} exists
- $Q \succeq P$ means $Q - P \succeq 0$

assume $X^TX \succ 0$ → $f(w) = y^Ty - 2w^TX^Ty + w^TX^TXw$

→ this function is Convex

→ Compute "derivative" & set to 0 to find minimizer

(Use gradients)

$$\nabla_{\underline{w}} f = \begin{bmatrix} \frac{df}{dw_1} \\ \frac{df}{dw_2} \\ \vdots \\ \frac{df}{dw_p} \end{bmatrix}$$

ex) $f(\underline{w}) = w^TQw$

$$= \sum_{i=1}^{p}\sum_{j=1}^{p} w_i Q_{ij} w_j$$

$$\frac{d(w_iQ_{ij}w_j)}{dw_k} = \begin{cases} 2Q_{ii}w_i & (k=i=j) \\ Q_{ij}w_j \text{ or } Q_{kj}w_j & (i=k\neq j) \\ Q_{ij}w_i \text{ or } Q_{ik}w_i & (i\neq k=j) \\ 0 & (k\neq i,j) \end{cases}$$

$\nabla_w f = Qw + Q^Tw$ if Q is symmetric

$\therefore \nabla_w f = 2Qw$

$f(\underline{w}) = y^Ty - 2w^TX^Ty + w^TX^TXw$

$\nabla_w f = 0 - 2X^Ty + 2X^TXw = 0$

$\Rightarrow X^TXw = X^Ty$ (X^TX is positive definite)

$\Rightarrow \hat{\underline{w}} = (X^TX)^{-1}X^Ty$