12. Orthogonal Projection Ⅱ

1) Orthogonal Projection and Linear Transformation

- Subspace W의 orthonormal basis $\{u_1, u_2\}$가 주어졌을 때, b의 orthogonal projection $\hat{b}$를 생각해보자.

$\hat{b} = f(b) = (b \cdot u_1)u_1 + (b \cdot u_2)u_2$ ⇒ u_1, u_2가 unit vector이므로 공식에서 변형됨.

$= (u_1^T b)u_1 + (u_2^T b)u_2$

↓ $u_1^T b$가 scalar이므로, 계산순서를 바꾸어도 결과는 동일.

$= u_1(u_1^T b) + u_2(u_2^T b)$

$= (u_1 u_1^T)b + (u_2 u_2^T)b$

$= (u_1 u_1^T + u_2 u_2^T)b$

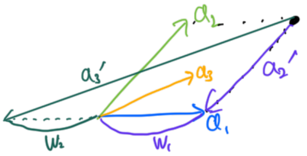
$= [u_1 \;\; u_2]\begin{bmatrix} u_1^T \\ u_2^T \end{bmatrix} b = UU^T b$ ⇒ linear transformation!

- $A = U = [u_1, u_2]$, $C = A^T A$가 invertible이면

$$\hat{b} = A\hat{x} = A(\underbrace{A^T A}_{I})^{-1}A^T b$$
$$= A(I)^{-1}A^T b$$
$$= AA^T b$$
$$= UU^T b$$

2) Projection & Regularization

- Matrix에서 두 column(feature)이 거의 linearly dependent라고 한다면, 그렇지 않을 때의 가중치와 차이가 많이 나게 된다. 즉, column의 linearly independent 여부에 따라 가중치가 민감하게 반응한다는 뜻이다.

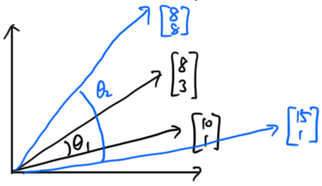


왼쪽 그림에서, a_2'은 (비록 orthogonal은 아니지만) a_1에 대한 projection으로 볼 수 있다.
그럼 가중치의 크기는 w_1 만큼이 된다.
→ dependency가 높아짐.
그런데 a_2가 아닌 a_3를 보자. 위의 경우와 동일하게 projection을 하면(a_3'),
가중치의 크기(w_2)가 매우 달라지는 것을 확인할 수 있다.

- Regularization 기법(Ridge, Lasso)은 두 column 사이의 각도를 벌리는 효과를 주어 column이 linearly independent에 가까워지도록 한다.

Ridge: $(A^T A + rI)^{-1} A^T b$

e.g. $\overset{A}{\begin{bmatrix} 10 & 8 \\ 1 & 3 \end{bmatrix}} + \overset{rI\ (r=5)}{\begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}} = \begin{bmatrix} 15 & 8 \\ 1 & 8 \end{bmatrix}$



$\theta_1 < \theta_2$