*기계학습 9주차 A4 1Page 정리. 2016104109 컴퓨터공학과 김병수

· Hold out Method

① 2개의 Subset으로 나눔
→ Train, Test

② Train set을 가지고 ML 모델 훈련시킴.

③ Test set을 가지고 Train 결과에 따른 성능 검증

· 3-Way Holdout Method.

① 3개의 Subset으로 나눔
→ Train, Test, Validation

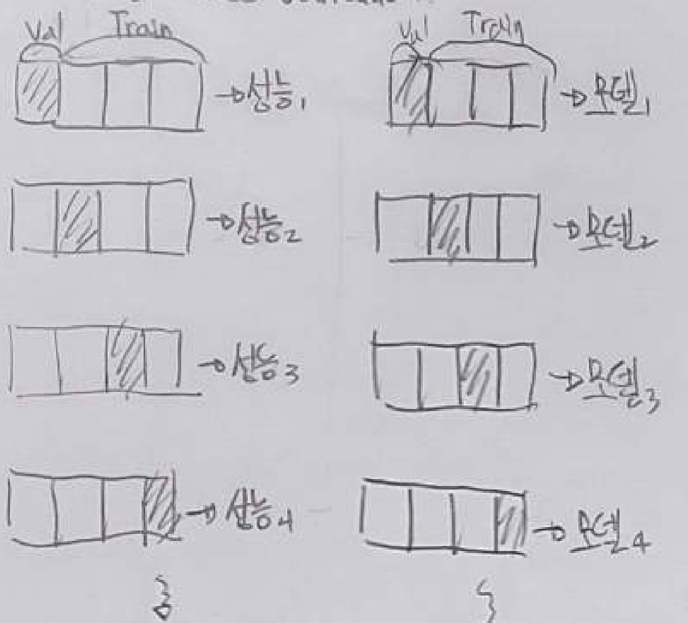
② Hyperparam Setting을 달리 하며,
Train Set을 가지고 ML 모델 훈련시킴

③ Validation Set을 가지고 일반화 성능검증

④ Test Set을 가지고 Train 결과에 따른 성능검증

· Cross Validation.

→ k-fold Cross Validation.

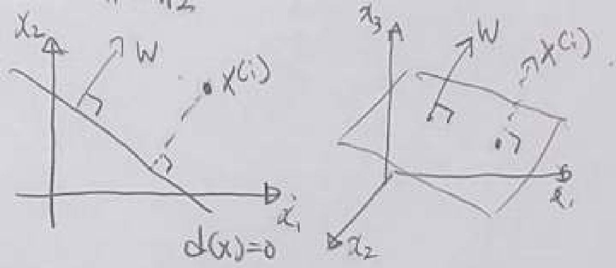


성능의 평균을 내어 성능 산출　　가장 좋은 성능 모델을 선택

* Decision Boundary의 Distance.

$d(x) = w^T x + b = 0 \quad (b = w_0)$.

$$h = \frac{|d(x^{(i)})|}{\|w\|_2}$$



ex) $d(x) = 2x_1 + x_2 - 4 = 0$. $x = [2, 4]^T$

$d(x) = w^T x + b = 0$ 에서, $w = [2, 1]^T$, $b = -4$.

$$\therefore h = \frac{|d(x)|}{\|w\|_2} = \frac{|2 \cdot 2 + 1 \cdot 4 - 4|}{\sqrt{2^2+1^2}} = \frac{4}{\sqrt{5}}$$

* Margin.

→ Decision Boundary에서. 양 class에 대한
Margin은 최대화할수록 일반화 능력 ↑

* SVM (Support Vector Machine)

→ Classification을 위한 ML Algorithm.

→ Margin을 이용하여 일반화 성능 ↑.

$$\text{Maximize } J(w) = \frac{2}{\|w\|_2} \quad (d(x) = \text{Const}).$$

Subject to $w^T x_i + b \geq 1, \ \forall y_i = 1$

$w^T x_i + b \leq -1, \ \forall y_i = -1$.