

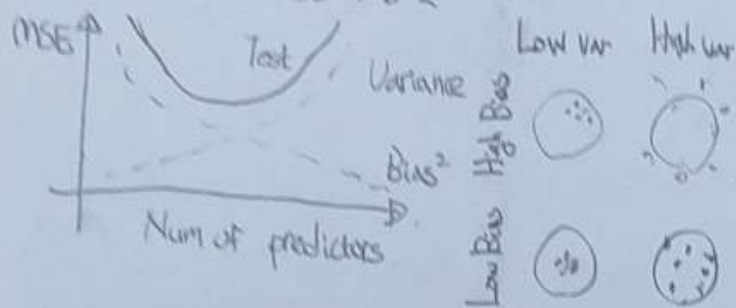
* Model Evaluation

Dataset: Training, Validation || Test
 ↓
 일반화 성능, model selection unseen data

* Prevent Overfitting

- ① 충분히 많은 Train Data 사용
- ② Validation Set을 이용한 Model Selection
- ③ Regularization을 사용하여 Model Param 규제

* 학습 모델 용량과 일반화 능력



$$Bias = E[e] = E[\hat{y}] - y$$

$$Var = E[(e - E[e])^2] = Var(\hat{y})$$

$$MSE(e) = E[e^2] = Var(\hat{y}) + Bias^2$$

$$Noise \triangleq E_x[(y(x) - \hat{y}(x))^2] = E_x[e^2] = \sigma_e^2$$

$\int_{-\infty}^{\infty} \hat{y}(x) + e$

* Regularization

- 명시적: 가중치 감소, 드롭아웃 등 신경망 구조나 목적함수를 직접 수정하는 방식
- 암시적: 조기 종료, 데이터 증대, 잡음 추가, 앙상블처럼 간접적으로 영향을 주는 방식

* Weight Penalty

- Regularized Cost Function

$$J_{\text{regularized}}(\theta; X, Y) = J(\theta; X, Y) + \lambda R(\theta)$$

규제 लागत된 목적함수 목적함수 규제함.

+ $R(\theta)$ 로 무엇을 사용할 것인지...? 가중치 정규화

$$L_2 \text{ norm: } R(\theta) = \|\theta\|_2^2$$

$$L_1 \text{ norm: } R(\theta) = \|\theta\|_1$$

* L_2 norm

$$J_{\text{reg}}(\theta; X, Y) = J(\theta; X, Y) + \lambda \|\theta\|_2^2$$

$$\therefore \nabla J_{\text{reg}}(\theta; X, Y) = \nabla J(\theta; X, Y) + 2\lambda \theta$$

$$\Rightarrow \theta = \theta - \rho \nabla J_{\text{reg}}(\theta; X, Y)$$

$$= \theta - \rho (\nabla J(\theta; X, Y) + 2\lambda \theta)$$

$$= (1 - 2\rho\lambda) \theta - \rho \nabla J(\theta; X, Y)$$

↳ θ 에 $(1 - 2\rho\lambda)$ 곱함 \Rightarrow 강제 미트하는 것

↳ 가중치를 작게 유지하여, 최종해는 원점쪽으로 당기는 효과

* L_1 norm

$$J_{\text{reg}}(\theta; X, Y) = J(\theta; X, Y) + \lambda \|\theta\|_1$$

$$\therefore \nabla J_{\text{reg}}(\theta; X, Y) = \nabla J(\theta; X, Y) + \lambda \text{sign}(\theta)$$

$$\Rightarrow \theta = \theta - \rho \nabla J_{\text{reg}}(\theta; X, Y)$$

$$= \theta - \rho (\nabla J(\theta; X, Y) + \lambda \text{sign}(\theta))$$

$$= \theta - \rho \nabla J(\theta; X, Y) - \rho \lambda \text{sign}(\theta)$$

↳ L_1 norm의 희소성 효과, 0이 되는 매개변수가 많은 선형 회귀에 적용하면 특징 선택 효과