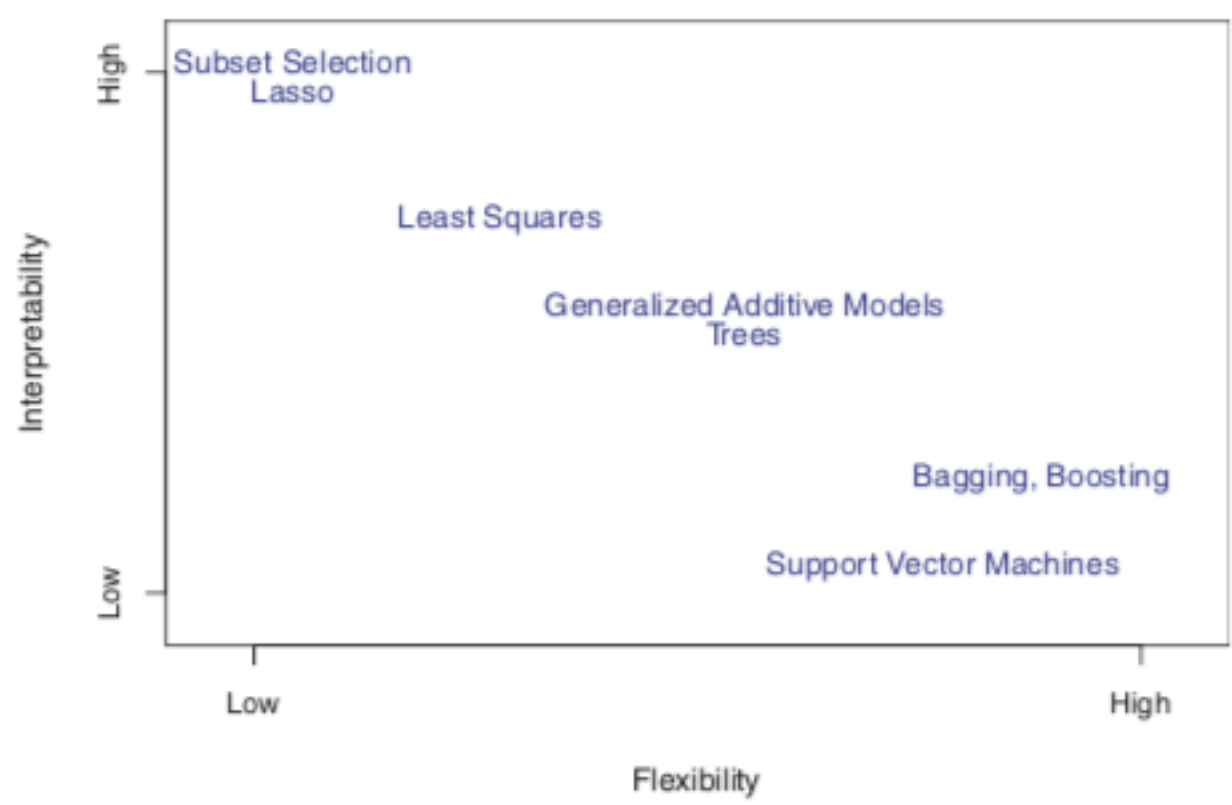


5. 모델 - 어떤 머신러닝 모델을 써야 할 것인가?

1) 목적에 따른 선택한다.

- Prediction - 가장 좋은 예측을 해야 한다.  
-> Flexiblity가 높은 모델을 사용한다.
- Inference: X, Y에 대한 관계에 대해서 이해를 하고 싶다. (물가가 상승하면 집값에 영향을 어떻게 미칠까)  
-> Interpretability가 높은 모델을 사용한다.



2) 성능이 좋은 모델을 선택한다.

- > Accuracy, Precision, Recall, ROC Curve 등을 바탕으로 모델을 선택한다.
- > Cross Validation 등의 방법을 사용한다.

cf) No Free Lunch Theorem

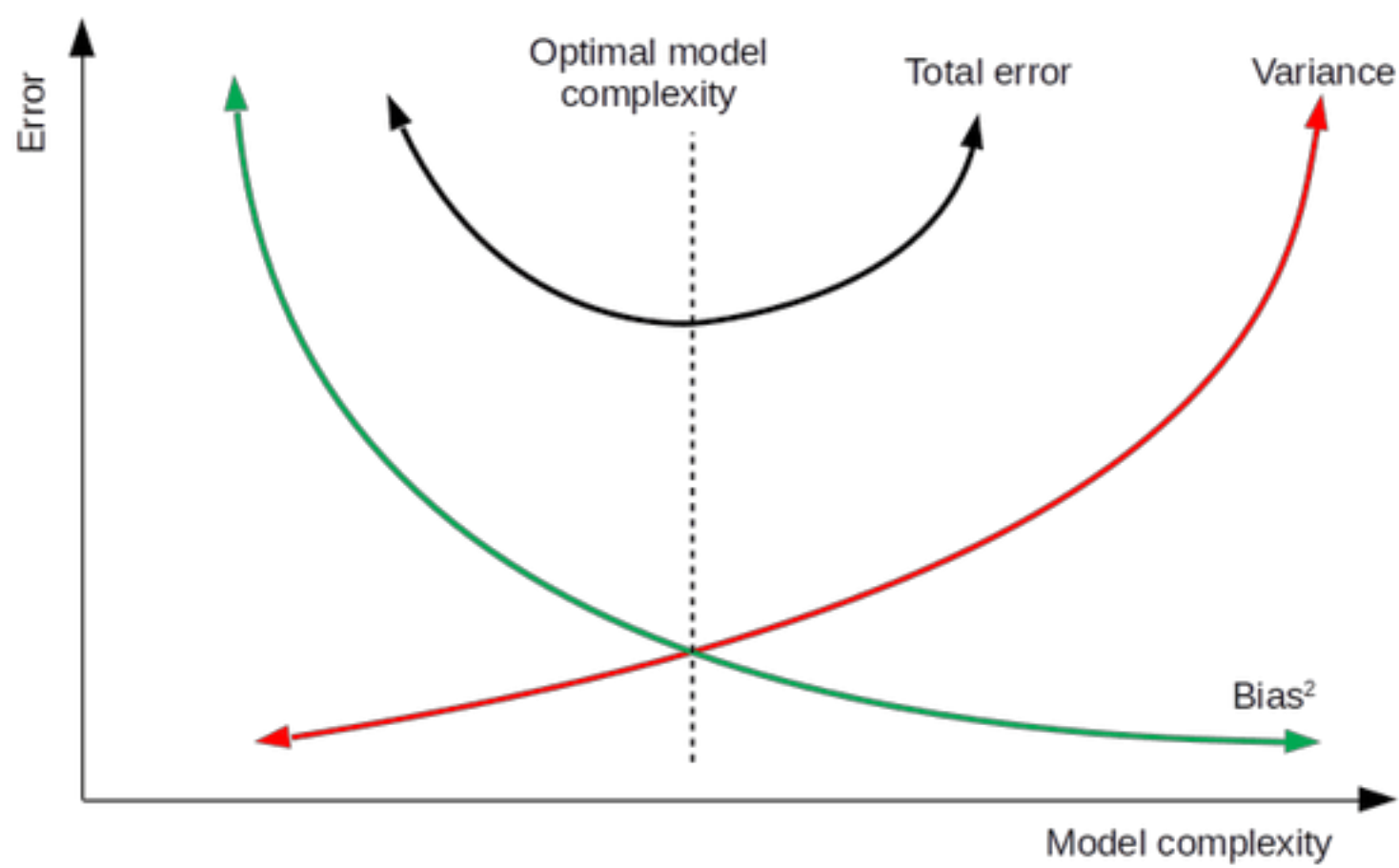
- > 모든 데이터에 대해서 가장 성능이 좋은 모델은 없다.
- > 데이터에 잘 맞는 모델을 선택한다.

3) 같은 성능이면 일반화가 잘되는 모델을 선택한다.

- > 같은 성능을 내는 모델이 있으면 간단한 모델일수록 좋다. (Ocam’s Razor)
- > overfitting의 위험을 줄여준다.
- > 모델을 간단하게 만드는 방법: Regularization

cf) Bias-Variance Tradeoff

- > 복잡한 모델은 Overfiiting이 되어 Bias가 적지만, Variance가 크다.
- > 간단한 모델은 Underfitting이 되어 Variance가 작지만, Bias가 크다.



이제

- 1) 통계와 머신러닝에서는 보통 모수 theta를 도입해서 특정 모수를 가정하는 모델(parametic method)로 추정을 시도한다.
- 2) theta에 대해서는 점추정을 할 수도 있고 분포 추정을 할 수도 있다.  
-> Frequentist: 점추정 + 구간추정 / Bayesian: 분포추정
- 3) machine learning의 개괄적인 분류  
-> Unsupervised, Supervised, Reinforcement Learning
- 4) 목적에 따라서 성능이 좋고 간단한 모델을 선택한다.

그렇다면 Machine Learning 모델을 실제로 어떻게 만들까?