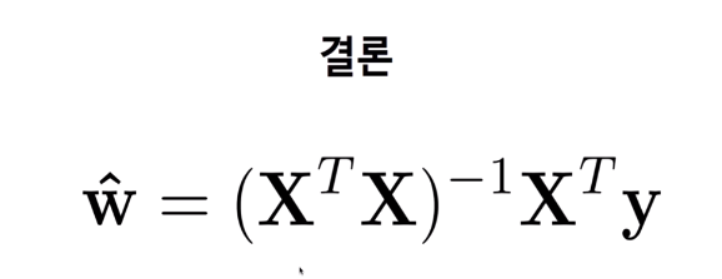
**통계 단기 정리**

1. **선형 회귀 (Linear Regression) -> Regression**

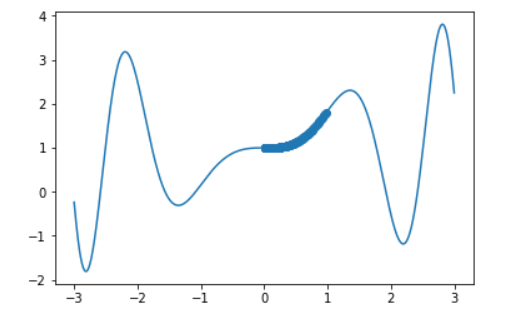
* Hypothesis 가설 즉 데이터가 이 모양을 갖추고 있을 것
* Cost function 비용함수 즉 최대한 줄이면 줄일수록 좋은 것!
* 모형 척도 (Performance)

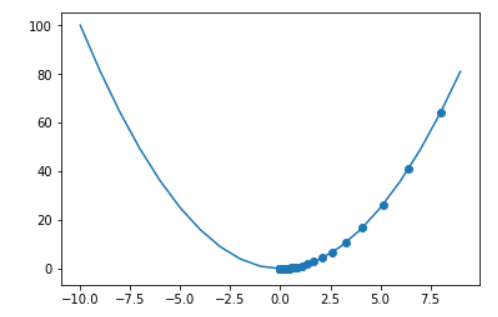
보통 연속형 - RMSE(Root Mean Square Error) 범주형 – Accuracy, F1 Score

* **정규방정식**(연립방정식의 해)

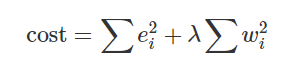


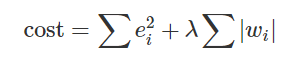
* **경사하강법** (Gradient Descent)

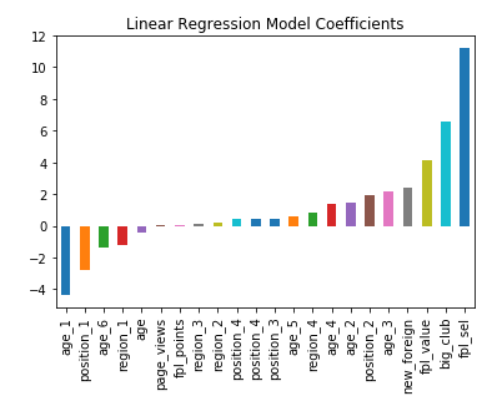
+ SGD(Stochastic Gradient Descent) … 한점이 아니라 여러점 씩 가는 것 확률적으로!

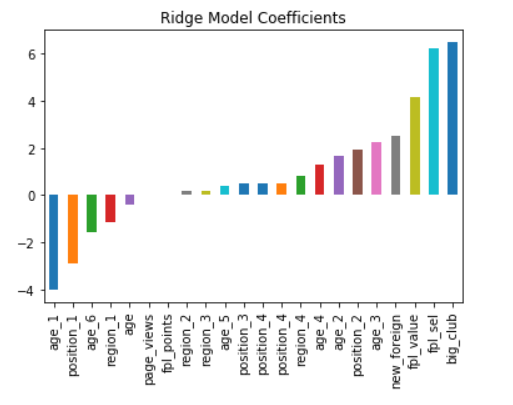
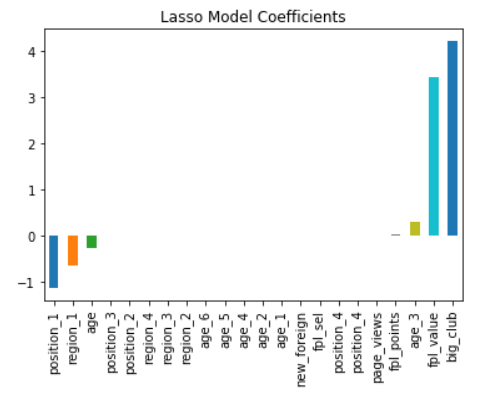


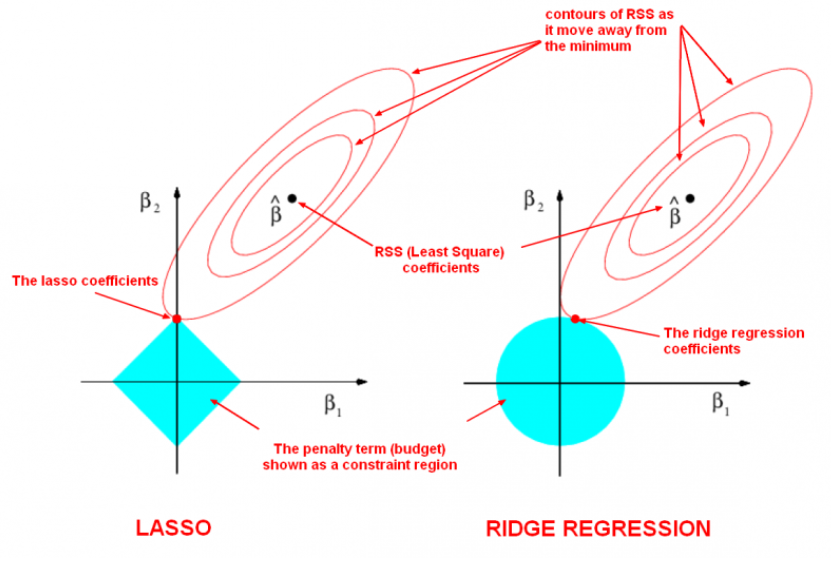
* **정규화 회귀**

왜?? Weight에 규제(Regularization)을 걸어 줌으로써 중요한 Feature를 더 중요하게!

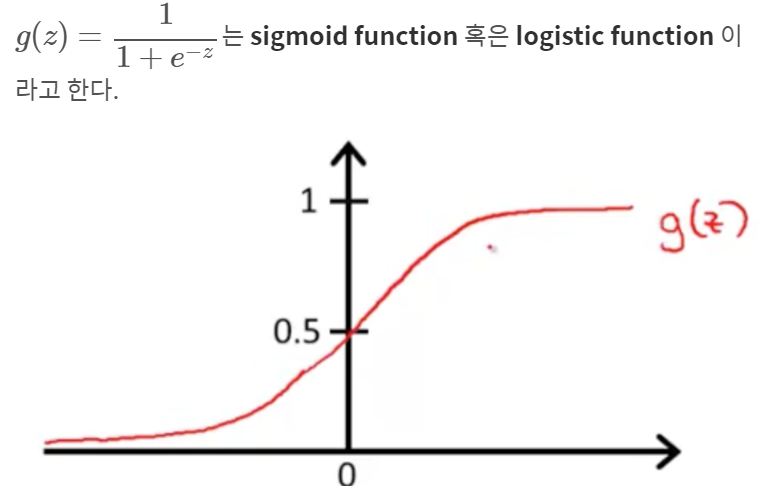
* + 릿지 회귀 (Ridge) L2 Regularization
  + 라쏘 회귀 (Lasso)
  + 엘라스틱 넷 회귀 (Elastic Net)





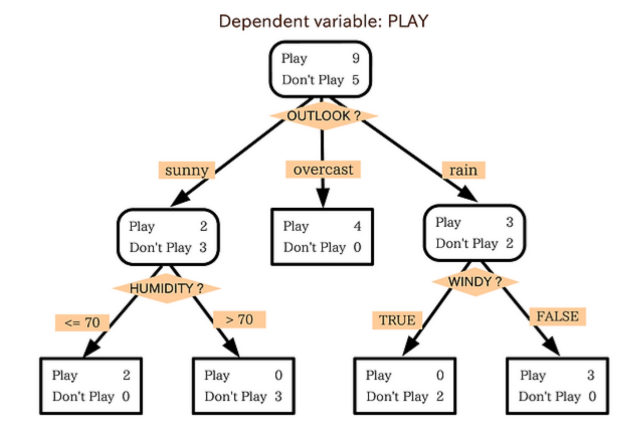


1. **로지스틱 회귀 (Logistic Regression) -> Classification**

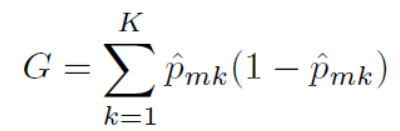
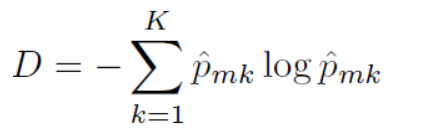


* 결국 이 아이도 GD 알고리즘 적용 가능

1. **의사결정나무 (Decision Tree)**



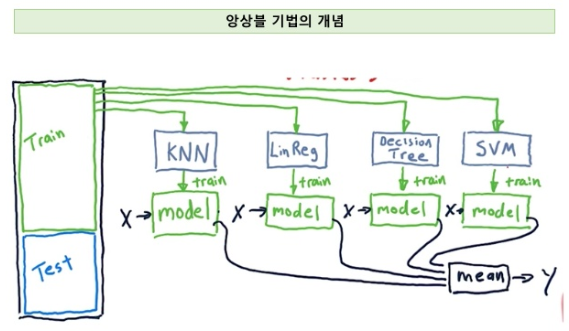
Cost Function으로 지니계수, cross-entropy



Overfitting을 막기 위해 가지치기(Pruning)를 함

**앙상블 (Ensenble)**

여러 모델을 사용해서 이 모델들을 평균을 취하는 방법 왜? Bias-Variance 관점



1. **배깅 (Bagging) … 트리를 앙상블기법(병렬 앙상블)을 사용한 것**

붓스트랩방식으로 트리를 겁나게 만든 다음에 그것들을 평균치하는 것

1. **랜덤포레스트 (RandomForest) 배깅의 일종**

배깅과는 달리 Tree간 correlation을 줄여준다.

How?? Random하게 Feature선택

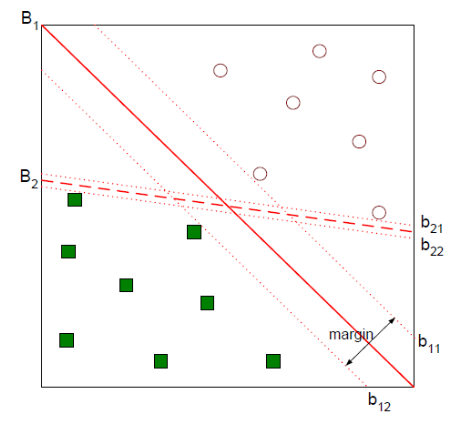
1. **부스팅 (Boosting) … 연속 앙상블(Error를 고려)**

배깅의 변형으로 오차(Error)에 가중치를 주어, 다음 모델이 학습하게 하는 알고리즘

Ex) GBM(Gradient Boosting Machine), AdaBoost(이건 잘 몰라용…)

XGboost - GBM의 속도와 성능면에서 향상시킨 알고리즘 … gbm + stochastic

1. **서포트 벡터 머신 (Support Vector Machine)**



쉽게 말하자면, 경계선과 가장 가까운 관측치를 제일 먼 선 or 공간(Plane)을 찾아주는 알고리즘!

그래서 b11과 b22사이를 마진(margin)이라고 함.

SVM은 이 마진을 최대화하는 분류 경계면을 찾는 기법

1. 주성분분석(Principal Component Analysis)
2. 인자분석(Factor Analysis)