

이 책에 나오는 대부분의 supervised learning algorithm은 확률분포추정이 기반.
MLE를 통해 vector theta의 값을 단순하게 할 수 있음.
위에 보이는 것은 linear regression. class 0, class 1 두 개 분류. Bernoulli func.
0과 1사이의 interval로 돼 있는 경우 == 이진 변수에 대한 분포
우리는 logistic sigmoid func으로 squash -> logistic regression.

SVM. 키워드는 아래와 같음.
margin: decision boundary와 support vector의 거리.
margin이 최대 -> $2/w$ 가 최대. -> w 가 최소

Robustness: 높을수록 Outlier에 영향을 받지 않음.

Kernel Trick: low dimensional space를 high dimensional space로 mapping

Radial Basis Function

C, gamma: RBF Kernel SVM에서 두 parameter C, gamma를 조정할 수 있음.
C: 커질수록 Outlier의 존재 가능성을 낮게 봄.
gamma: Gaussian Function의 표준편차와 관련되어 있음.
gamma가 클수록 표준편차는 작아짐.
하나의 data sample이 영향력을 행사하는 거리를 결정.
커질수록 한 data point들이 영향력을 행사하는 거리가 짧아짐.

Grid Search: 두 parameter C, gamma는 경험적인 방법으로 선택함.
두 parameter의 여러 combination을 테스트해서 가장 좋은 performance를 내는 것을 택함.

k-Nearest Neighbors (kNN)

장점

기존 분류 체계를 이용하므로 정확도 높음
가까운 k개 데이터만 활용 -> 오류 데이터는 비교대상 제외
기존 데이터가 기반 -> 데이터 대한 가정이 없음

단점

기존의 모든 데이터를 비교 -> 데이터 많을수록 처리시간 증가
데이터 활용 많음 -> 고용량 메모리 사용
curse of dimensionality를 피하기 위해 dimensionality reduction

Decision tree는 두 가지 data type으로 나뉜다: [Classification], Regression.

Decision tree 생성을 위해 반복적인 분할로 트리를 학습해야 함.

분할 기준은 부모마디보다 자식마디의 순수도가 증가.

반복시행한 후 모든 공간을 직사각형으로 나누어서 각 직사각형이 가능한 순수하고 동질적이 되도록.