

기계학습 1주차 과제

12191656 이채연

HomeWorks #1

Supervised learning(지도 학습)과 unsupervised learning(비지도 학습)의 차이점에 대해 기술하십시오.

1. Supervised Learning 문제와 Unsupervised Learning 문제는 Training Dataset D 를 define할 때 차이점이 있다. Supervised Learning에서는 training sample에서 x_i 라는 feature vector와 함께 y_i 라는 레이블이 있지만, Unsupervised Learning에서는 레이블이 없고 x_i 라는 feature vector만 있다.
2. Model의 form
 - Supervised learning : $p(y_i | x_i, \theta)$
 - Unsupervised learning : $p(x_i | \theta)$ θ 는 학습해야될 model의 parameter.
Supervised Learning은 $p(y_i | x_i, \theta)$ 로 표현할 수 있다. Response variable, label을 estimation하는 문제이다.
Unsupervised learning에서는 label을 예측하는 것이 아니라 output variable이 없기 때문에, 데이터 자체에 대한 distribution, density를 찾아내는 문제이다.
3. SL problem에서는 y_i 가 single variable이다. 즉 most cases에 univariate probability estimation 문제이다. Target value y_i 를 사용하기 때문에 좀 더 제한적인, assumption을 하고 문제를 푸는 것이기 때문에 좀 더 정확하게 문제를 해결할 수 있다. 하지만 USL problem에서는 x_i 가 multivariate이다. 즉 multivariate 문제이다. Unsupervised learning에서는 label을 예측하는 것이 아니라 output variable이 없기 때문에, 데이터 자체에 대한 distribution, density를 찾아내는 문제이다. 그래서 unsupervised learning이 어렵다.
4. Supervised learning은 conditional density estimation이지만 unsupervised learning은 unconditional density estimation이다.
5. 또한, Unsupervised Learning 문제는 Supervised Learning 문제와 비교했을 때 잘 정의되어있지 않은 문제(less well-defined problem)이다. Dataset이 주어질 때 어떤 종류의 pattern이 거기에 있는지, 어떤 metric을 사용해서 이것을 분류해야 될지

를 찾아내야 하는 문제이기 때문에 Supervised Learning 문제와 비교했을 때 알고리즘을 설계하기가 어렵고 까다롭다.

HomeWorks #2

기계학습에서 확률 모델이 필요한 이유에 대해 설명 하시오.

모델이 classification을 할 때 어디에 속했다고 100% 단정할 수 없는 상황인 경우, 확률적으로 예측을 해야한다. 확률적이라는 것은 classification에 대한 믿음이라고 생각할 수 있다. 그래서 input vector x 와 training set D 가 주어졌을 때 우리는 그것에 대한 가능한 label에 대한 probability로 확률적으로 표현할 수 있다. 그리고 여러 개의 model을 가지고 예측을 하는 경우, input vector x 와 training set D 와 함께 M 이라는 모델 파라미터를 포함시켜서 모델 중 하나가 주어졌을 때 y 를 예측하는 문제를 정의할 수 있다.