

기계학습 7주차 과제

기상데이터 이용 태양광 발전량 예측

컴퓨터공학과
17011584 정재경

개요

주어진 데이터셋은 시간별 기상정보가 나타나 있어, 각 정보에 따른 태양광 발전량을 나타낸다. 제공된 정보에 따라서 태양광 발전량을 예측하는 문제로, 회귀문제에 해당한다.
이번 과제는 Random Forest 방식을 사용해서 해결해보았다.

실험

이번 과제에서 변경을 해보며 결과값의 차이 추이를 본 하이퍼파라미터는 n_estimators, max_features 이다.

n_estimators	max_features	Score
100	auto	0.99854
100	sqrt	0.99235
100	log2	0.99580
1000	auto	0.99865
1000	sqrt	0.99275
1000	log2	0.99609

결론

이 데이터셋은 n_estimators = 1000, max_features = auto 일 때의 성능이 가장 좋았다. Ensemble 방식의 학습은 확실히 이전에 배웠던 기법들 보다 학습시간이 많이 늘어난 것을 체감할 수 있었다. 특히나 n_estimators 는 Random Forest에서 Tree의 갯수를 의미하는 하이퍼파라미터 인데, 100과 1000의 학습 시간 차이가 유의 미하게 나는 것을 확인할 수 있었다.