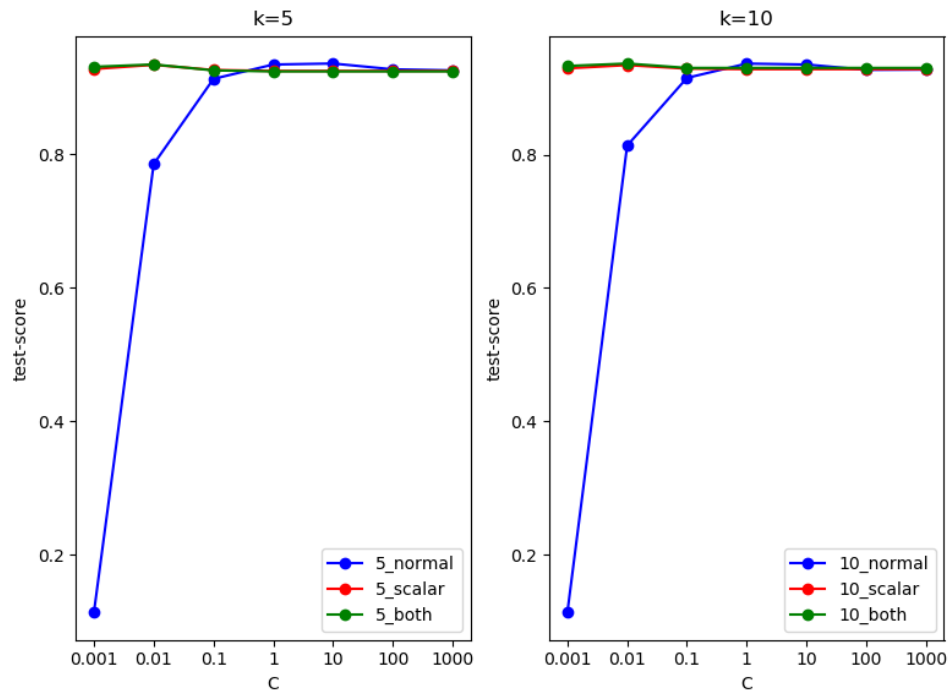


결과물 설명

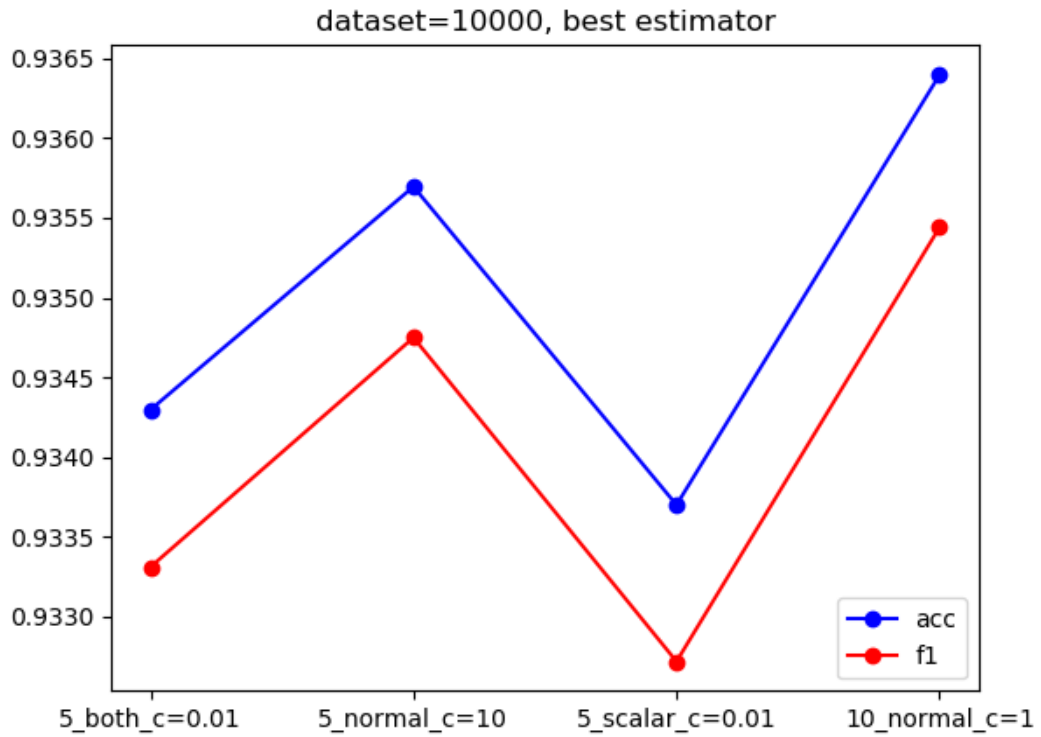
1. dataset = 10000개, testset = 60000개

전처리기로 [Normalizer, StandardScaler, 두개 다]를 사용했을 때의 최적 C 값을 찾아봄
이 때 k-fold cross validation을 사용했는데, k=5, k=10인 경우 총 2*3=6번의 CV를 함.



최적 C

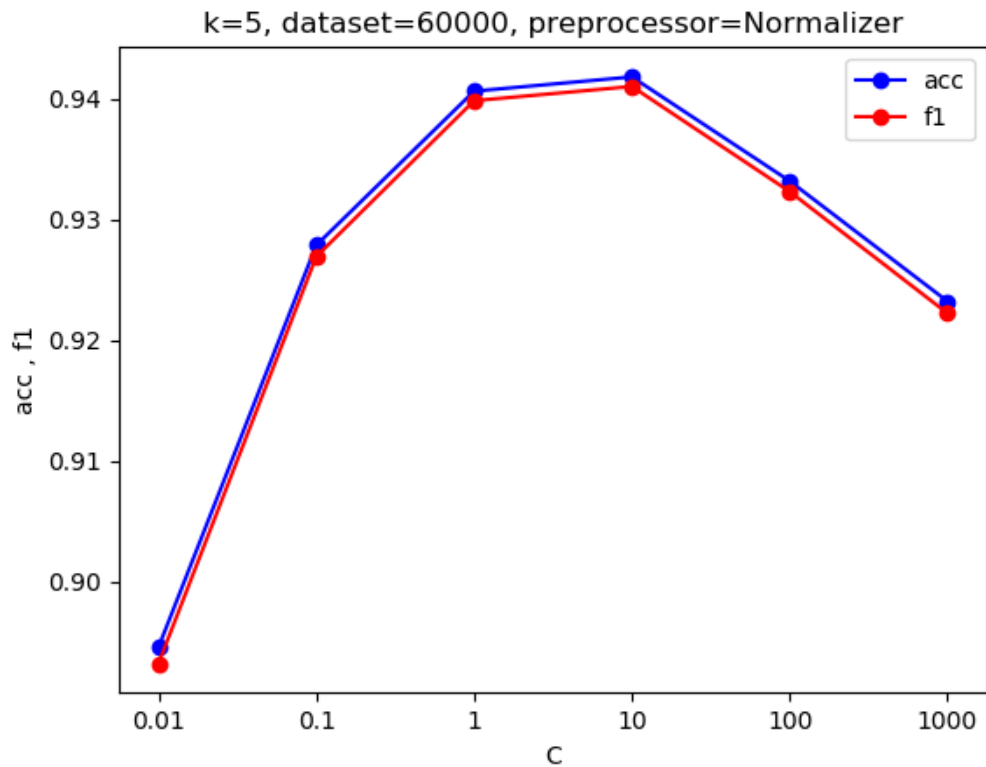
- scalar, both : C= 0.01
- normal : k=5일 때 C=10, k=10일 때 C=1
- 위 4가지 Cross Validation 결과 (accuracy, f1-score) 비교



테스트 결과

60000개의 testset에 대해 test 한 결과 4가지 모두 f1-score가 0.92가 나왔다. (10개 클래스별 f1-score 평균)

2. dataset = 60000개, testset = 10000개



60000개의 경우 학습에 오래걸리기 때문에 k=5개로 고정하고 그 중 가장 잘 나왔던 Normalizer를 전처리기로 사용했을 때에 최적 C값을 찾아보았다.

모델의 정확도는 $C=10$ 일 때 가장 높았다.

테스트 결과

Testset 10000개에 대해 C 에 대해 테스트를 한 결과 $C=1$ 과 $C=10$ 일 때가 0.95로 가장 높게 나왔다.

