

Chapter 7 앙상블 학습과 랜덤 포레스트

7.1 앙상블 방법은 예측기가 가능한 한 서로 (독립적/의존적)일 때, 최고의 성능을 발휘한다.

답: 독립적, 다양한 분류기를 얻는 한 가지 방법은 각기 다른 알고리즘을 학습시키는 것으로 이렇게 하면 매우 다른 종류의 오차를 만들 가능성이 높아 앙상블 모델의 정확도를 향상시킨다.

7.2 BaggingClassifier에서 페이스팅을 사용하는 방법은?

답: bootstrap=False로 바꾼다.

7.3 BaggingClassifier에서 특성 샘플링을 지원하는데, 이 때 샘플링은 두 매개변수로 조절됩니다. 그 두 매개변수는 무엇인가요?

답: max_features, bootstrap_features.

7.4 랜덤포레스트 알고리즘은 어떻게 동작하는지 설명하세요.

답: 트리의 노드를 분할할 때, 전체 특성 중에서 최선의 특성을 찾는 대신 무작위로 선택한 특성 후보 중에서 최적의 특성을 찾는 식으로 무작위성을 더 주입한다. 이는 결국 트리를 더욱 다양하게 만들고 편향을 손해보는 대신 분산을 낮추어 전체적으로 더 훌륭한 모델을 만들어 낸다.

7.5 부스팅이란 무엇인가요?

답: 약한 학습기를 여러 개 연결하여 강한 학습기를 만드는 앙상블 방법이다.

7.6 블렌더를 학습시키는 일반적인 방법으로 한가지를 들자면?

답: 홀드 아웃 세트 사용.