

4주차 과제

2023170832 고지원

1. 수백만 개의 특성을 가진 훈련 세트에서는 어떤 선형 회귀 알고리즘을 사용할 수 있는가?
 - 확률적 경사 하강법, 미니배치 경사 하강법
2. 훈련세트에 있는 특성들이 각기 아주 다른 스케일을 가지고 있을때, 이런 데이터에 잘 작동하지 않는 알고리즘은 무엇인가? 그 이유는 무엇이고, 이 문제를 어떻게 해결할 수 있는가?
 - K-최근접 이웃 알고리즘(KNN), 이유 : 거리계산왜곡, 해결 : 정규화, 표준화
3. 경사 하강법으로 로지스틱 회귀 모델을 훈련시킬 때 지역 최솟값에 갇힐 가능성 이 있는가?
 - 없음. 전역 최솟값으로 수렴
4. 충분히 오랫동안 실행하면 모든 경사 하강법 알고리즘이 같은 모델을 만들 것인가?
 - 손실 함수의 특성과 하이퍼파라미터, 초기값에 따라 달라짐.
5. 검증 오차가 상승하면 미니배치 경사 하강법을 즉시 중단하는 것이 좋은가?
 - 과대 적합 방지를 위해 조기 중단 기법 사용.
6. 어떤 경사 하강법 알고리즘이 가장 빠르게 최적 솔루션의 주변에 도달하는가?
 - 확률적 경사하강법
7. 다항회귀를 사용했을때 학습 곡선을 보니 훈련 오차와 검증 오차 사이에 간격이 크다. 이때 이 문제는 무엇인가? 해결 방법에는 무엇이 있는가?
 - 원인 : 과적합 / 해결 : 모델 단순화, 데이터 증대, 교차 검증
8. 릿지회귀를 사용했을때 훈련 오차와 검증 오차가 거의 비슷하고 둘 다 높았습니다. 이 모델에는 높은 편향이 문제인가? 아니면 높은 분산이 문제인가? 이때, 규제 하이퍼파라미터 알파값을 어떻게 해야하는가?
 - 높은 편향 문제. 알파값을 줄여야함.
9. 사진을 낮과 밤, 실내와 실외로 분류하려고 할때, 두 개의 로지스틱 회귀 분류기를 만들어야 하는가? 하나의 소프트맥스 회귀 분류기를 만들어야 하는가?
 - 하나의 소프트맥스 회귀 분류기를 사용하는 것이 더 효율적. 관리와 예측 용이, 모든 클래스 동시에 고려해 학습 가능