

머신러닝

11week 모델 정리

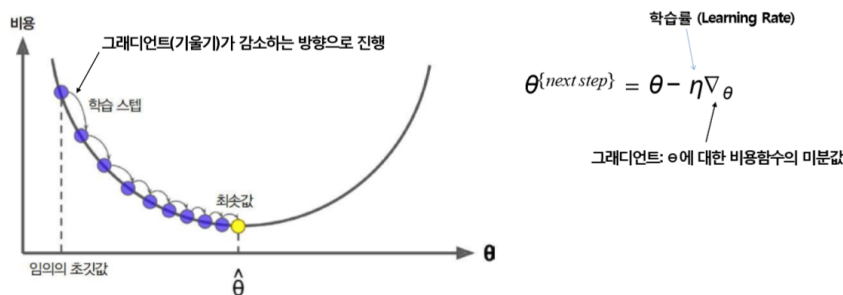
이연희

1. **선형 회귀:** 직접 계산할 수 있는 공식을 사용해서 훈련 세트에 가장 잘 맞는 모델 파라미터를 해석적으로 구한다. 경사 하강법(Gradient Descent)이라 불리는 최적화 방법을 사용하여 모델 파라미터를 조금씩 업데이트 하면서 비용 함수(loss)를 훈련 세트에 대해 최소화시킨다. 결국 이전에 했던 방식과 동일한 파라미터로 수렴한다.

MSE (Mean Squared Error): 위의 선형 회귀 모델을 학습시켜 Train Set에 잘 맞는 파라미터를 찾아야 한다. 이러한 파라미터를 찾은 뒤 잘 맞는지 성능을 측정하는 지표로는 평균 제곱근 오차(RMSE, Root Mean Squared Error)가 있다. RMSE는 MSE식에 루트를 씌운 것이므로, 결국 MSE를 최소화하는 파라미터 W 를 찾아야 한다.

- 정규방정식: 비용 함수를 최소화하는 θ 를 찾기 위한 해석적인 방법(수학 공식)이다.

2. **경사 하강법:** 경사 하강법에서 중요한 하이퍼파라미터는 학습률(learning rate)이다. 학습률이 너무 작으면 수렴하는데까지 시간이 오래 걸리고, 학습률이 너무 크면 발산하게 된다. 경사 하강법을 사용할 때는 반드시 모든 특성(feature)들이 같은 스케일을 가지도록 해야 한다 (Standard 또는 MinMax 등). 그렇지 않으면 학습시간이 오래 걸리게 된다



- **확률적 경사 하강법:** 경사 하강법의 가장 큰 문제는 매 스텝(step)에서 전체 Train Set을 사용해 Gradient Descent를 계산한다는 것이다. 따라서, Train Set가 커지면 학습이 매우 느려 지게 된다. 이를 해결하기 위해, 확률적 경사 하강법(Stochastic Gradient Descent)은 매 스텝에서 랜덤하게 하나의 데이터(샘플)을 선택해 Gradient Descent(GD)를 계산한다. 따라서, 전체 데이터를 이용해 GD를 계산하는 것보다 속도는 빠르지만 훨씬 불안정하다. 비용 함수(Loss Function)가 최소값에 수렴할 때까지 부드럽게 감소하지 않고, 위아래로 요동치면서 평균적으로 감소한다. 이처럼 비용 함수가 불규칙하게 요동치면서 감소할 경

우 지역 최소값(local minimum)을 건너뛸 수 있는 가능성이 있기 때문에, SGD가 전역 최소값(global minimum)을 찾을 가능성이 높다. 하지만 이러한 무작위성은 지역 최소값을 탈출할 수 있지만, 전역 최소값에는 다다르지 못하는 경우가 있다. 이를 해결하기 위해 학습률(learning rate)을 점진적으로 감소시키는 learning rate decay 기법을 사용한다. learning rate decay(또는 learning rate schedule)는 학습을 시작할 때는 학습률을 크게 하고, 점진적으로 학습률을 줄여 전역 최소값에 도달하게 하는 방법이다.

- **배치 경사 하강법:** 경사 하강법을 구현하려면 모델 파라미터에 θ_j 대해 비용 함수의 기울기를 계산해야 한다. 다시 말해 θ_j 가 조금 변경될 때 비용 함수가 얼마나 바뀌는지 계산해야 한다. 이것을 **편도 함수**라고 한다.
 - **미니배치 경사 하강법:** 미니배치 경사 하강법(Mini-batch Gradient Descent)은 각 스텝에서 전체 Train Set을 미니배치(mini-batch), 즉 작은 데이터셋을 추출한 뒤 Gradient를 계산하는 방법이다. 미니배치 경사하강법의 장점은 행렬 연산에 최적화된 하드웨어, GPU에서 빠르게 수행되는 것이다. 미니배치 경사 하강법은 SGD에 비해 덜 불규칙하게 감소하지만, local minimum에 빠질 확률은 높은 경우가 있다.
3. **다항 회귀 (Polynomial Regression):** 다항 회귀(polynomial regression)는 각 특성(feature)의 제곱을 새로운 특성으로 추가하여 선형 모델을 학습시키는 방법을 말한다.
4. **학습 곡선 (Learning Curve):** 고차 다항 회귀를 사용하면 선형 회귀(linear regression)보다 Train Set에 Overfitting 될 위험이 크다. 규제(Regularization)가 있는 선형 모델: 과대적합을 감소시키는 좋은 방법. 모델을 규제하는 것. 자유도를 줄이면 데이터에 과대적합되기 더 어려워진다. 예를 들어, 다항 회귀 모델을 규제하는 간단한 방법은 다항식의 차수를 감소시키는 것. 선형 회귀 모델에서는 보통 모델의 가중치를 제한함으로써 규제를 가한다. 가중치를 제한하는 방법은 **릿지 회귀**, **라쏘 회귀**, **엘라스틱넷**이 있다.
- **릿지 회귀 (Ridge Regression):** 학습 알고리즘을 데이터에 맞추는 것뿐만 아니라 모델의 가중치가 가능한 작게 유지되도록 노력한다. 규제항은 훈련하는 동안 비용 함수에 추가된다. 모델의 훈련이 끝나면 모델의 성능은 규제가 없는 성능 지표로 평가한다. α 가 커지면 가중치가 0에 가까워지고, α 가 0에 가까우면 MSE 만 남는다. 릿지 회귀는 입력 특성의 스케일에 민감하기 때문에 수행하기 전에 데이터의 스케일을 맞추는 것이 중요하다.

미분 결과를 간단하게 만들기 위해 추가

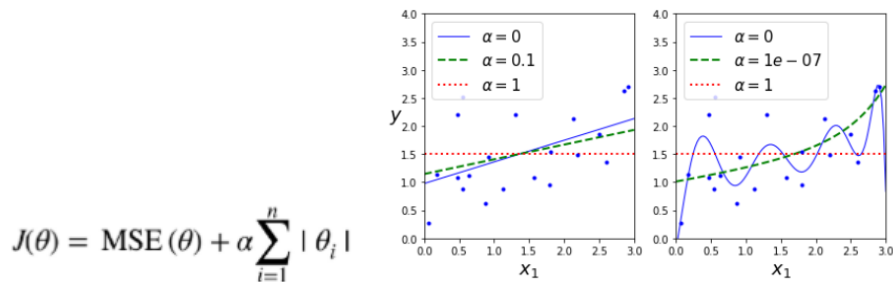
규제량을 조절: 하이퍼파라미터

$$J(\theta) = \text{MSE}(\theta) + \alpha \frac{1}{2} \sum_{i=1}^n \theta_i^2$$

θ_i 는 규제하지 않음

- **라쏘 회귀 (Lasso Regression):** 선형 회귀의 또 다른 규제된 버전. 릿지 회귀처럼 비용 함수에 규제항을 더하지만 아래와 같이 사용한다. 덜 중요한 특성의 가중치를 완전히 제

거하려고 한다. 예를 들어, 오른쪽 그래프에서 점선이 2차 방정식처럼 보이고 선형이다. 차수가 높은 다항 특성의 가중치가 모두 0이 되었다. 다시 말해서 라쏘 회귀는 자동으로 특성 선택을 하고 희소 모델을 만든다. 릿지 회귀는 규제가 강해지면 함수가 부드러운 선의 형태로 그려지지만 차수는 유지된다. 라쏘 회귀는 규제가 강해지면 다항식에서 일부 항이 없어지고 차수가 낮아지며 특징 선택으로 사용할 수 있다.



$$J(\theta) = \text{MSE}(\theta) + \alpha \sum_{i=1}^n |\theta_i|$$

- **엘라스틱넷 (Elastic Net):** 엘라스틱 넷(Elastic Net)은 릿지(Ridge) 회귀와 라쏘(Lasso) 회귀를 절충한 모델이다. 규제항은 릿지와 회귀의 규제항을 단순히 더해서 사용하며, 혼합 정도는 혼합 비율 r 을 사용해 조절한다. $r=0$ 이면 엘라스틱넷은 릿지 회귀와 같고, $r=1$ 이면 라쏘 회귀와 같다.

✓ **선형회귀, 릿지, 라쏘, 엘라스틱넷을 언제 사용하는가?** 규제가 약간이라도 있는 것이 대부분의 경우에 좋으므로 일반적으로 평범한 선형 회귀는 피해야한다. 릿지가 기본이 되지만 실제로 쓰이는 특성이 몇 개뿐이라고 의심되면 라쏘나 엘라스틱넷이 낫다. 이 모델들은 이전에 이야기한 것처럼 불필요한 특성의 가중치를 0으로 만든다. 특성 수가 훈련 샘플 수보다 많거나 특성 몇 개가 강하게 연관되어 있을 때는 엘라스틱넷을 사용한다

- **조기 종료:** 경사 하강법과 같은 반복적인 알고리즘을 규제하는 색다른 방식. 검증 에러가 최소값에 도달하면 바로 훈련을 중지.

5. **로지스틱 회귀(Logistic Regression):** 분류에서도 사용할 수 있는 회귀 알고리즘. 샘플이 특정 클래스에 속할 확률을 추정하는데 사용. 선형 방정식을 시그모이드 함수에 통과시켜서 0 ~ 1 사이의 확률을 계산.

- **확률 추정:** 로지스틱 회귀는 선형 회귀(linear regression)와 같이 입력 특성(feature)의 가중치 합을 계산한 뒤 로지스틱 함수(sigmoid)를 적용해 출력 값을 계산한다.
- **하나의 샘플에 대한 비용 함수(확률 p , 양성/음성 클래스 y)**
- **전체 훈련 세트에 대한 비용 함수:** 해당 비용 함수는 최소값을 계산하는 알려진 해가 없다. 비용 함수가 다행히 볼록 함수 이므로 경사 하강법을 사용해 전역 최소값을 찾는 것을 보장한다.
- **결정 경계:** 붓꽃 데이터셋 사용. 3개의 클래스(setosa, virginica, versicolor). 4개의 특징

(꽃잎(petal)의 폭과 너비, 꽃받침(sepal)의 폭과 너비). 각 클래스당 150개의 샘플

6. **소프트맥스 회귀 (Softmax Regression):** 로지스틱 회귀 모델을 여러 개의 이진 분류기를 훈련시켜 연결하지 않고 직접 다중 클래스를 지원하도록 일반화. 회귀 모델이 각 클래스에 대한 점수를 계산하고, 그 점수에 소프트맥스 함수를 적용하여 각 클래스의 확률을 추정. 비용 함수는 크로스-엔트로피 함수를 사용. 두 개의 클래스만 존재한다면($K=2$), 이 비용 함수는 로지스틱 회귀의 비용 함수와 같다.