

모델공부

ENCORE@JUPYTER

RNN

이전 단계의 출력을 현재 단계의 입력으로 쓸수있다면 큰 RNN덩어리를 만들어서 다른 모델의 INPUT으로 넣을수있는가?

기울기 소실이 발생한다. 역전파를 쓸때에는 합성곱미분을 쓴다

=> 활성화 함수와 관련

SIGOMID라고 계속 함수를 사용하면 0일때 기울기가 켈 높지만 계속 곱해지면 $0.25^{\text{제곱}}$ 씩으로 거의 모든값이 0으로 수렴한다.

=>이걸 막기위해서 나온게 TANH인데 값을 2배로 늘렸기에 기울기가 0일때 1값을 뛰지만 이것도 역시 TANH이므로 x값이 작거나 크게되면 SIGMOID처럼 의미가 없게된다.

=> 그러므로 RELU가 등장했는데 0이하에서는 0이고 양수이면 무조건 1이므로

input x => output x 그대로 돌려준다.그리고 연산이 쉽다. 미분해야할필요가 없고 그대로 x값이므로

다만 입력값이 음수인뉴런은??

근데 음수값이 음수일수가있는가?? 정규화에 따라 달리진다.

=> LEAKY RELU가 등장했는데 음수일때 0.1이나 0.01로 하는데 이게 의미가 잇는가?

조금이라도 소실을 막는다.

모델 스테킹 OR 모델 합성이라고 한다.

모델이 복잡해지면 과적합이 커지게된다.

=> 출력층 직전 은닉층 노드수 줄이기, 설명변수가 되기에

output이 뒤로만 가면 포워드 신경망이라고 한다

나중에 양방향 신경망도 있다.

one to many

이미지 캡셔닝 사진인식하면 제목출력

many to one

감정분류, 감정분석, 스팸메일 분류, 주로 분류에 쓰이는듯

many to many

데이터 정규화

항상 모델전에 정규화를 스케일링을 하지만 layer를 거칠수록, 출력값이 데이터분포가 바뀐다.

=>이래서 저번에 배치정규화를 넣어줘야했구나 그래서 예러가 생긴듯...

이러한 현상을 ics라고한다.

손실함수와 떨어진 layers수록 가중치를 높게 잡는다 무시될수있으므

$$BN(X) = \gamma \left(\frac{X - \mu_{batch}}{\sigma_{batch}} \right) + \beta$$

- X : 입력 데이터
- γ : 추가 스케일링
- β : 편향
- μ_{batch} : 배치별 평균값($\frac{1}{B} \sum_i x_i$)
- σ_{batch} : 배치별 표준편차($\frac{1}{B} \sum_i (x_i - \mu_{batch})^2$)

r,b는 선형을 바꾼다. 학습단계시 배치스케일링인데, 이걸 역전파 알고리즘에 따라 가능할수있다 왜냐하면 선형이므로...

모델평가

배치

몇개의 문제를 풀고 매길것인가??

iteration

총 batch 수이며 step이라고도 부른다.

dataset 크기 / 배치 크기 = iteration크기

epoch

총 문제집 몇번풀래?

오차함수

mae - 절대값, 중앙값에 영향을 많이 받아 이상치 무시 가능, 회귀에 주로 사용

배치경사하강법

한번의 epoch 이고 batch1개 batch_size는 데이터개수이다.

확률적 경사하강법

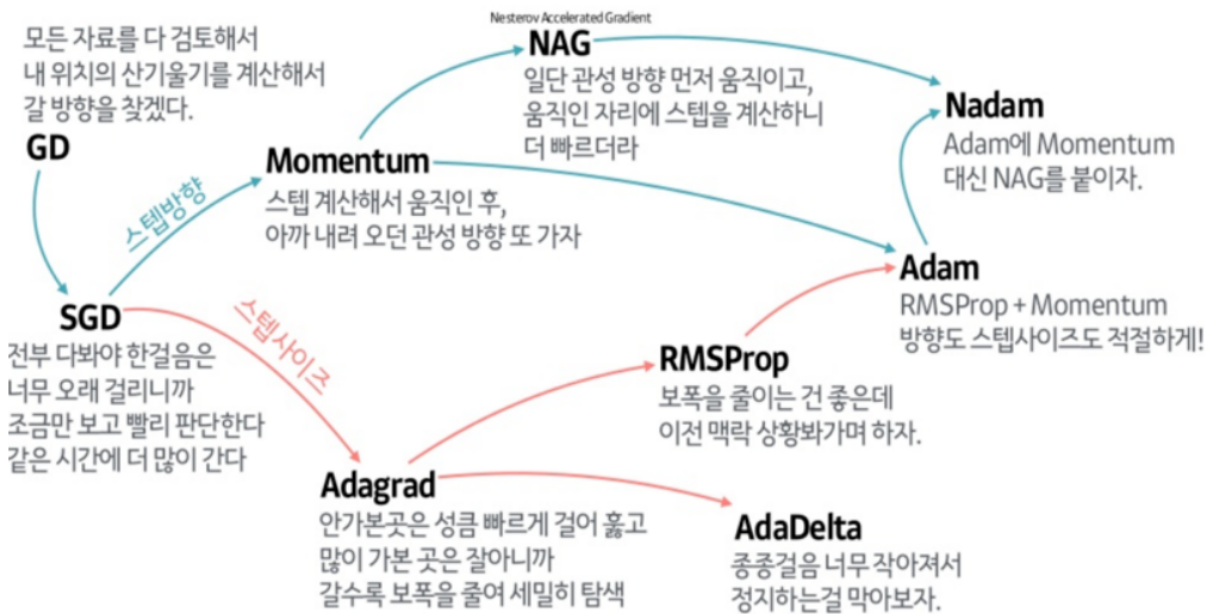
한번에 epoch하는 배치경사하강법이 시간을 오래걸리므로 batch_size를 1로 해서 하지만 불안정하므로 정확도가 낮게된다..

미니배치경사하강법

두개의 종합통합본 주로 batch_size는 2의 n제곱..

학습률

최적의 가중치를 찾아준다. 다만 학습률 조절에 따라정확도는 올라가지만 오래걸린다.



nag는 미리 앞을 본다

경사하강법

1. 블록 함수-말그대로 위아래로 볼록 누구나 찾을수있다. 저점을 찾기쉬
2. 비블록함수 - 어디가 최저점이고 모른다, 고등학생대배운 삼차함수 미분...

momentum

이전의 기울기를 기억하여 그 기울기만큼 추가로 이동, 물리학법

adagrade

변수마다 학습률을 다르게 조절 좋아보인다.

학습이 잘되면 학습률 낮추고 그 반대는 올리는데 무조건 양의 무한대로 가면 값이 0에 수렴하게된다.

$$g_t = g_{t-1} + (\nabla f(x_{t-1}))^2$$

$$x_t = x_{t-1} - \frac{\eta}{\sqrt{g_t + \epsilon}} \cdot \nabla f(x_{t-1})$$

후반에 갈수록 $g_t + \epsilon$ 가 커지므로 학습이 잘되서 근사해지는건가 아니면 추가적으로 학습이 되지않는건가 헷갈림.

rmsprop은

$$g_t = \gamma g_{t-1} + (1 - \gamma)(\nabla f(x_{t-1}))^2$$

$$x_t = x_{t-1} - \frac{\eta}{\sqrt{g_t + \epsilon}} \cdot \nabla f(x_{t-1})$$

기울기를 업데이트한다. 지수이동평균은 최근값에는 가중치를 많이두고 시점이 멀어질수록 가중치를 낮게둔다 약간 lstm그런거랑 유사한듯
무한대로 발산하는 걸 방지하기위해서 adamgrad처럼 작은값 ϵ 를 더해준다.

adam

momentum과 rmsprop을 결합한방식

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla f(x_{t-1})$$

$$g_t = \beta_2 g_{t-1} + (1 - \beta_2)(\nabla f(x_{t-1}))^2$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \hat{g}_t = \frac{g_t}{1 - \beta_2^t}$$

$$x_t = x_{t-1} - \frac{\eta}{\sqrt{\hat{g}_t + \epsilon}} \cdot \hat{m}_t$$

- β_1 : Momentum의 지수이동평균 ≈ 0.9
- β_2 : RMSProp의 지수이동평균 ≈ 0.999
- \hat{m}, \hat{g} : 학습 초기 시 m_t, g_t 가 0이 되는 것을 방지하기 위한 보정 값
- ϵ : 분모가 0이 되는 것을 방지하기 위한 작은 값 $\approx 10^{-8}$
- η : 학습률 ≈ 0.001

편향을 보정한것이다. 초기값이 0에 가까워지는걸 막기위해

퍼셉트론

입력값*가중치+편향 = > 출력

단층 퍼셉트론 - xor gate 구현못함

2차원 평면으로만 xor gate 구현못함 즉 층이 1개가 더 필요함

다중 퍼셉트론

분석절차

1. 데이터불러오기
2. 불필요한 변수제거
3. 결측치 처리
4. 이상치 확인 시각화 및 제거
5. 전처리 끝 데이터 저장
6. 데이터 병합
7. x,y값 시각화
8. 데이터 스케일링
9. 학습/테스트 분리
10. 모델링(하이퍼파라미터 모델 성능 시각화 함수를 쓰자)<https://heytech.tistory.com/150>

과대적합

train loss와 valid loss 의 교차지점

=>early stopping

교차검증

특정데이터셋에만 적용이 되지않고 일반화가 가능하다,

순서가 있는 시계열 데이터라도 shuffle하고 사용가능

CNN

POOLING 레이어가 있어야 크기줄임

max average 다양한 기법이 있는 average가 좋아보이나 average는 smothng 시켜서 cnn의 장점을 못살린다.

Data Augmentation

데이터가 양이 부족할때 데이터를 늘리는 거손 아주 좋은 방법이다. 그 기법들은 다양하다.
진짜 신박하다

전이학습

prophet 모델

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t.$$

비주기적변화, 주기적인변화, 불규칙이벤트, 오차항
매개변수를 빨리 바꿀수있다. 결측치 있어도 오케이

$g(t)$

상한하한이 존재하는 모델

$$g(t) = \frac{C}{1 + \exp(-k(t - m))},$$

하지만 C가 수용가능상수인데, 상수가 아닌경우도 존재한다.....

k는 성장률.

저걸 막기위하여서

$$g(t) = \frac{C(t)}{1 + \exp(-(k + \mathbf{a}(t)^T \boldsymbol{\delta})(t - (m + \mathbf{a}(t)^T \boldsymbol{\gamma})))}.$$

C가 시간에 영향을 받을수도 있고, 어떠한 기준을 이벤트를 기준으로 성장률이 바뀔수도있다.

상한하한이 없는 모델

piecewise linear 모델

$$g(t) = (k + \mathbf{a}(t)^\top \boldsymbol{\delta})t + (m + \mathbf{a}(t)^\top \boldsymbol{\gamma}),$$

로직스틱의 지수부분을 가져옴.

여기서 어떠한 기준으로 바뀌는 시점은 자동탐지가 된다.

$s(t)$

계절성

$$s(t) = \sum_{n=1}^N \left(a_n \cos \left(\frac{2\pi nt}{P} \right) + b_n \sin \left(\frac{2\pi nt}{P} \right) \right)$$

P는 정규주기 일주일이면 7

연간계절성에 따라사라서는 n이 3이나 10이좋다.

$h(t)$ 이벤트 사건성

이건 주기적패턴을 안뒀다.

holidays 패키지에 이미 한국 휴일들이 들어가있다.

```
model = Prophet.Prophet(changepoint_prior_scale=10, changepoints = ['2020-01-25','2020-02-29'],yearly_seasonality=10,weekly_seasonality=False,daily_seasonality=False, growth='linear')
```

각각 년월일 계절성 추가가능

knn

위도경도가 비슷한 애들끼리 클러스터링해서 mean값으로 온도값을 예측해 가중치 30

그리고 그 국가에 따른 선형회귀모델을 따로만들어서 lstm모델을 가중치 70

두 값을 결합

answer_온도가 있으므로 가장 mse값이 작은값을 기준으로 잡는다.