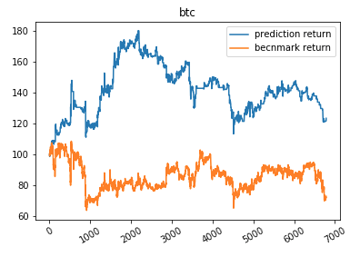
# 1. [실습] Simple RNN Model for Crypto Currency Prediction 실습을 따라해 봅시다 (Day1)\*

아마 이번 실습을 진행하시면서 다들 의아해 하셨을 만한 부분이 있는데요

" 왜 이렇게 수익이 안나오지? " 라는 생각을 하셨을 것 같습니다

저도 동일한 질문을 가지고 있고, 미국의 경우에는 가장 efficient한 마켓이다 보니, 가격에 데이터에만 기반한 알고리즘을 이용해서는 수익이 나는 알고리즘을 찾기가 쉽지않은것 같습니다. 다음주 부터는 한국 데이터에 대한 모델을 짜기 시작하고요. 완전히 동일한 모델에 mix max scale만 추가하여도 아래와 같은 performance가 나옵니다



물론 슬리피지와 거래 수수료를 고려하지 않은 모형이여서 실제 수익으 더 낮을 것이고요, 그래도 다음주 부터는 수익이 나는 알고리즘을 짜면서 더 재밌게 실습을 진행하실 수 있지 않을까 싶습니다

1번 질문으로 돌아가서 답변을 드리자면

## 질문: normalized scale 중 중간 중간에 shuffle을 하는데, sequence data이다보니 처음에는 shuffle은 안해야 하는게 정상이 아닌가 하는 생각을 해봤습니다.

해당 질문에 답변을 드리자면, shuffle은 모든 sequence를 생성한 이후에 일어나기 떄문에 투자 시기에 관계없이 performance가 나오도록 하는 모델의 일반화에 도움이 됩니다.

그런데 만약에 투자 시기 또한 중요한 요인이라 local time에 대한 모델이 맞다고 생각하시면 shuffle 자체가 적절하지 않을수도 있습니다. 이 부분은 추가적인 테스트가 필요한 부분이고요

## 질문: deque() 동작은 이해하겠는데 왜 해야하는지 잘 모르겠습니다.

deque는 maxlen을 설정하여 time 기준의 sequence 데이터를 생성할 수 있는데 간편한 구조를 제공합니다.

저희가 만드는 데이터의 모양이

……..

……..

……..

……..

……..

이런식으로 한칸식 밀리면서 과거 60개 포인트로 미래 시점 데이터 예측하는데, 다음 sequence와 이전 sequence 사이에는 60개 포인트 중에서 겹치는 부분이 59개이고 나머지는 하나씩만 새로 넣고 과거꺼는 없애는 형식으로 만들게 됩니다.

이런 데이터를 생성하는데, deque만큼 깔끔한 구조가 없는것 같고요

혹시 해당 부분에서 조금 더 설명이 필요한 부분이 있으신가요~?

## 질문: buy data와 sell data 수량을 동일하게 가져갔습니다. 한쪽으로 학습이 치우치는걸 방지하기 위함이라고 생각이 드는데, 각각 3천개가 넘는 data인데도, 충분히 많은데도 이런 부분을 신경써야하는지요?

양쪽 label (buy, sell)을 동일하게 가져가는건 모델이 safe answer (하나의 답만 하는)를 방지하는데 도움이 됩니다. 지금은 문제가 양호한 편이지만, 예를들어서 80번 오르고 20번 내린 경우에 모델이 항상 오른다고만 예측해도 예측률이 80% 이기 때문에, 양쪽을 동등하게 해주어야 safe answer를 보다 잘 피할 수 있게 됩니다.

여기까지가 문제 1번에 질문이여서 다음 2번 질문으로 넘어가겠습니다

# \*2. [과제] RNN Model을 수정하여 본인만의 Time Series Prediction Model을 작성하여 봅시다 (Day 1)\*

해당 질문과 직접 관련하여서는 질문이 없었어서 추가 질문 중에 하나를 다루겠습니다

## 질문: 두개의 다른 정보를 다른 모델로 해서 학습해서 하나로 적용시킬때 어떤식으로 하는건지?? 예를 들어서 코인의 가격변화를 통해 결과를 알려주는 모델과 뉴스를 통해 결과를 알려주는 모델이 있다고 할때 두 모델을 함께 적용시키고 싶을때 어떤식으로 하면 되나요? (edited)

질문 주신 모델은, 두개의 정보를 쓰는 좋은 모델인데요!

기존에 학습하신 model을 load 하신 뒤에, 각각의 layer를 새롭게 받아, 새로운 model을 정의하실 수 있습니다.

정말 간단하게 sudo 코드를 짜봤는데

\* model\_a = Model(inputs\_a, outputs\_a)

\* model\_b = Model(inputs\_a, outputs\_b)

\* new\_inputs = concatenate([outputs\_a, outputs\_b])

\* new\_outputs = Dense(~~~~~)(new\_inputs) # or more deep architecture

\* model\_c = Model(new\_inputs, new\_outputs)

이런식으로 model a의 모양이 위와 같이 생긴경우에, 각각의 output을 concat하여 새로운 모델의 입력으로 주시면 됩니다. 그러면 다음 질문으로 넘어가겠습니다

# \*3. [과제] 직접 개발한 Time Series Prediction Model을 평가하여 봅시다 (Day 2)\*

개인적으로는 day2에서의 자료가 제가 생각하는 저희 수업에서 가장 original한 자료가 아니였을까 싶습니다 ㅎㅎ

어떤 논문에서나 블로그 등 어디에서도 다뤄주지 않는 내용인데, 금융의 background가 아니면 사실 잘 알기 어려운 부분에 대해서 다루었습니다.

해당 자료에서도 질문이 있었는데 질문: Risk까지 반영한 수익을 지속적으로 낼 수 있는 모델이 설계가 가능할까요?

제가 언급드린 measure들은 custom loss function을 추가하여 만드실 수 있습니다.

예를 들자면, 각 batch를 단위로 return의 상위 10% 하위 10%를 truncate하여 나머지 데이터에 기반한 geometric return에 마이너스를 붙힌 값을 loss로 주게되면 첫번째 measure에 대한 custom loss function이 됩니다. 이런식으로 나머지 아이디어에 대해서도 모델에 반영해 주실 수 있습니다

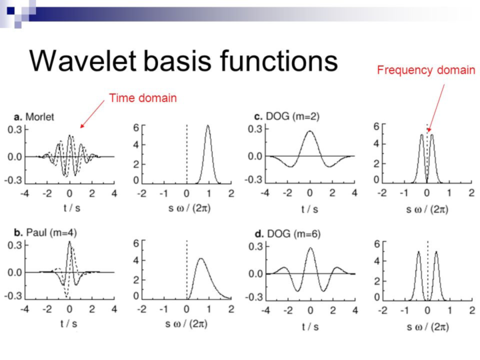
# \*4. [과제] Time Series Prediction Model의 Performance를 Wavelets Denoise를 통해 개선시켜 봅시다 (Day 3)\*

wavelets에 대해서는 정말 질문이 많았는데요. wavelets는 쉬운 내용이 아닌 것 같습니다

## 질문: Denoising에도 여러 방법이 있을것으로 생각 되는데 wavelet을 사용하시는 이유가 있나요?

여기서 부터는 사실 답이 있다기 보다는, 혼자서 했으면 어렵게 느꼇을 만한 내용을 시도 해본 것이고요. 그래도 해당 denoising 방법의 특징을 말씀드리자면 frequency와 time domain을 동시에 고려하는 모델인데 이러한 방법론에는 크게 1. 푸리에 변환 2. 웨이블랫 변환이 있습니다.

그런데 1번의 경우에는 sin function으로 원레의 signal을 모사하는데 local pattern을 찾기에 sin function이란게 적절하지 않습니다. 때문에 wavelets의 basis 특히나, admissibility condition이라고 해서



이런 느낌의 아래 위 합이 0인 (필요조건)을 만족하는 basis들을 잡아서 local pattern들을 찾게 됩니다

푸리에에서 쓰는 sin과는 다르게, sin은 한곳의 패턴을 매칭하려고 가져온 basis가 훨씬 멀리있는 곳에도 영향을 미치기 때문에 문제가 있는데, wavelets는 그런 부분이 없습니다.

조금 더 설명을 드리자면, wavelets는 어떤 signal을 다른 함수 (basis)들의 조합으로 근사하는데요.

근사에 사용된 특정 basis에 대해 여러 scale을 적용하여 basis set을 만들게 됩니다. 이러한 basis set에 기반해서 원레의 signal과 매칭되는 부분 / 매칭 되지 않는 부분을 쪼개면 level1 filter가 되고요 (threshold에 대한 criterion이 필요하긴함) 매칭 되지 않은 부분에 대해서 반복적으로 basis set에 기반해서 매칭된 부분 / 매칭되지 않은 부분을 쪼개나가면 level2, level3 등이 됩니다

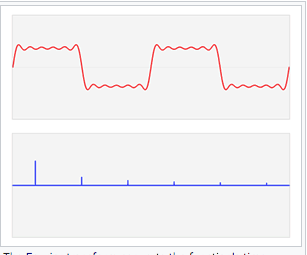
저희는 level1에만 기반하긴했는데 사실, threshold에 민감해서 어떤 값을 쓰는게 최적일지는 추가적인 실험이 필요할 것 같습니다.

## 고재민(10월딥러닝일요10시) [5 minutes ago] 노이즈를 줄이는데는 간단하게는 MA도 있을수 있고, FIR fileter가 있을 수도 있는데 wavelet만의 장점이 있을까요?

ma는 time domain은 있어도 frequency domain이라고 할만한 부분이 없을것 같습니다. FIR filter는 어떤것인지 정확히는 모르지만 filter들의 핵심은 보통 trend 파트와 cycle 파트를 잘 나눠보고 어떤 filter가 trend만을 잘 뽑아내는가? 에 대한 답변이 되어야 하는것 같습니다.

이 부분은 제가 느끼기에도 정답이 있다기 보다는 실험을 통해서 알수 있는 부분인것 같고요

고재민(10월딥러닝일요10시) [2 minutes ago] 네 알겠습니다



위의 어떤 패턴이 있는 signal을 frequency로 나타내면 훨씬 간단하게 표현할 수 있습니다. 조금 더 자세히 wavelets의 장점을 설명드리자면..

위에서도 언급하였듯이, 이런 방법은 2개가 있는데 fourier는 sin으로 하다보니

local pattern을 매칭하는데는 적절하지 않은 부분이 있습니다

해당 부분에 대해 혹시 추가적인 질문이 없으시면 다음 질문으로 넘어가겠습니다 :slightly\_smiling\_face:

# \*5. [과제] Time Series Prediction Model의 Performance를 다양한 Hyperparameter실험을 통해 개선시켜 봅시다 (Day 4)\*

6 replies

김종호(스터디진행자) [1 minute ago]

해당 자료가 사실은 앤드류 옹 4주 강의를 하나의 자료로 요약한 자료인데요

김종호(스터디진행자) [< 1 minute ago]

급하게 만든 부분이 종종 있어서, 혹시 설명이 부족한 부분은 추가로 질문을 해주시면 됩니다~

김종호(스터디진행자) [< 1 minute ago]

질문: 슬라이드 L2 norm penalty makes weight vector to be sparse => Simpler model에 대해 설명 좀 부탁드립니다.

김종호(스터디진행자) [< 1 minute ago]

여기에 답변을 드리자면, 사실 저 멘트는 lasso ridge regression을 염두해 두고 적은 내용인데

김종호(스터디진행자) [< 1 minute ago]

나중에 보니 설명이 빠졌더라고요 ㅠ

나중에 보니 설명이 빠졌더라고요 ㅠ

김종호(스터디진행자) [< 1 minute ago]

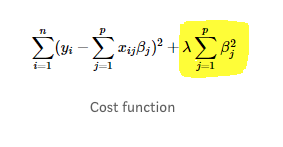
핵심만 전달드리자면,

김종호(스터디진행자) [< 1 minute ago]

특정한 weight가 큰 경우에 loss를 높히는 패널티를 줄수 있습니다

김종호(스터디진행자) [< 1 minute ago]

식으로 보자면



이 부분의 노란색으로 칠한 부분 인데요

추정한 계수에 제곱을 하면, L2 norm (제곱)을 계산하게 되는데, weight의 크기에 따라서 선형적인 패널티를 주는 것이 아니라 제곱 함수를 따라서 기여하게됩니다.

예를 들어서 x 와, 2x가 있으면 x는 제곱해봐야 x^2인데 2x는 제곱하면 4x^2으로 값이 큰 애들이 L2 norm을 패널티로 보면 값이 작은 애들에 비해 훨씬 큰 패널티로 작용을 합니다.

따라서 L2 norm penalty를 추가하면 큰 weight 값에 제약을 걸어서 모델을 단순하게 만들어 줍니다.

쉽게는 link의 weight을 0으로 만들면 네트워크에 연결된 link의 총 개수가 줄어드니까 더 간단한 모델이 된다고 이해하셔도 좋고요.

혹시 해당 PT에서 추가적인 질문이 있으신가요~?

# \*이번주에 학습하시면서 이해가 안가셨거나 궁금하신 질문들을 모두 적어주세요.\*

김종호(스터디진행자) [5 minutes ago]

이번 주는 스터디를 어떻게 진행하셨나요~?

박제권 (10월딥러닝일요10시) [4 minutes ago]

이번주는 소스이해하기에 급급한 정도였습니다. ^^;

## 최명환(10월딥러닝일요10시) crypto 실습시 input shape을 잠시 생각해봤는데요. 마지막 feature가 8인데, 2개 buy or sell로 생각해보면 2개로도 고민해 볼수 있지 않을까 하는 생각을 해봤습니다. 아직도 input shape에 대해서 고민스러운것 같긴합니다.ㅎㅎ;

김종호(스터디진행자) [3 minutes ago]

one hot encoding을 하시면, 얘기주신 모델이 되는것 같습니다. 신기한게 같은 모델일것 같은 것도 binary classification 문제로 푸는지 one hot으로 2개를 넣어서 푸는지에 따라 모델 성능이 달라진다고도 하는것 같습니다.

확인해 보지는 못했지만, one hot으로 하는게 더 좋다라는 말도 있고요 ㅎㅎ

최명환(10월딥러닝일요10시) [2 minutes ago]

네

김종호(스터디진행자) [2 minutes ago]

이번주에는 정말 많은 주제를 다루기도 했고, 사실 다른 강의에서는 한달에 걸쳐 배우는걸 압축적으로 진행한 부분도있는것 같습니다

스터디를 추가로 하시다가 나중에라도 궁금하신 부분이 있으시면 물어봐주셔도 좋고요

최명환(10월딥러닝일요10시) 혹 시간이 되시면 2day에서 return 모델로 하신게 있는데 전체적인 맥락 설명을 부탁드려되 될까요? 코드를 봐도. 첫번째 예제와 어떤 차이로

김종호(스터디진행자) [1 minute ago]

기존과 다른 부분은 딱 2개가 있습니다

1. return을 만드는 것 (x(t)-x(t-1)/x(t-1))

2. return을 예측하는 것 (continuous value)

기존과는 다르게 continuous value에 대한 예측이다보니 loss function이 cross categorical entropy가 아니

mae를 쓰고있습니다

혹시 해당 부분에서 또 궁금하신 부분이 있으신가요~?

최명환(10월딥러닝일요10시) 네. 감사합니다. 살펴보겠습니다. ㅎㅎ