

HAI_STOCK 레퍼런스

🕒 작성일시	@2022년 8월 5일 오전 11:09
🕒 최종 편집일시	@2022년 8월 19일 오후 12:50
📄 문서 유형	hai stock
🔗 레퍼런스	

알파 팩터 리서치

알파 팩터는 자산 가격 움직임을 예측하는 것을 목표로 하는 원시 데이터의 변환이다. 자산 수익률을 높이는 리스크를 포착하게 설계된다. (12장에서 비지도 학습을 이용해 데이터 기반의 합성 팩터를 도출하는 방법을 알아본다.)

모멘텀 팩터

모멘텀 팩터는 추세와 패턴을 식별함으로써 전통적으로 가격 시계열의 변화에서 도출된다. 이는 여러 기간에서 자산을 횡단면 적으로 비교하거나 자산의 시계열을 분석함으로써 절대적 수익률이나 상대적 수익률을 기반으로 구축될 수 있다.

흔히 사용되는 예시적 지표들을 다음과 같다.

- 상대 강도 지수, RSI : RSI 는 최근 가격 변화를 종목 간 비교해 과매수 또는 과매도 되는 종목을 확인한다.
- 가격 모멘텀 : 과거 거래일수에 대한 총수익률을 계산한다.
- 가격 가속도 : 장기와 단기(1년과 3개월)에 대한 일일 가격에 대한 선형 회귀식을 이용해 추세의 기울기를 계산하고 가격 가속도의 척도로 기울기의 변화를 계산한다.

변동성과 규모 이례 현상

저변동성 팩터는 평균 이하의 변동성, 베타 또는 특이 리스크로 주식의 초과 수익률을 포착한다.

다음은 저변동성을 식별하는데 사용되는 척도에 대한 설명이다.

- 우량주 팩터
우량주 팩터는 시장 대비 우량한 기업에 대한 초과 수익을 포착하는 것을 목적으로 한다.

칼만 필터를 사용한 알파 팩터의 잡음 제거

칼만 필터는 시계열과 같은 순차적 데이터에 대한 동적 선형 모델로, 새로운 정보가 도달하면 이에 대해 적응하는 모델이다. 이동 평균과 같은 고정 크기의 윈도우 또는 지수 이동 평균과 같은 주어진 가중치 세트를 사용하는 것이 아니라 확률 모델을 기반으로 새로운 데이터를 시계열의 현재 값 추정에 반영한다.

신호에서 트레이딩으로: 백테스트를 위한 zipline

여기서는 과거 평균으로부터 최근의 성과가 얼마나 벗어나 있는지를 측정하는 단순 평균 회귀 팩터를 개발하고 테스트할 것이다.

단기반전(short-term reversal) 은 1분 이하에서 1달까지의 기간에 걸쳐 롤링 평균으로 주가가 회귀하는 경향의 약한 예측 패턴을 이용하는 일반적인 전략이다.

정보 계수

ML 은 예측 목적을 최적화하는 것이며, 이번 절에서 알파 팩터의 성과를 측정하는 주요 척도를 소개한다. 알파를 벤치마크를 초과하는 수익률의 평균으로 정의한다.

알파 팩터의 목적은 미래 수익률의 정확한 방향성 예측이다. 따라서 자연스러운 성과 척도는 알파 팩터 예측과 타깃 종목들의 미래 수익률과의 상관관계다.

트레이딩을 위한 ML: 기본

6. 머신러닝 프로세스

커버할 모델의 범주는 다음을 포함한다.

- 횡단면, 시계열과 패널 데이터의 회귀와 분류를 위한 선형 모델
- 의사결정 트리와 같은 비선형 트리 기반의 모델을 포함한 generalized additive models
- 랜덤 포레스트와 그레디언트 부스팅 머신을 포함하는 앙상블 모델
- 차원 축소와 클러스터링을 위한 비지도 선형 및 비선형 방법
- 순환 신경망과 합성곱 신경망
- 강화학습 모델

7. 선형모델: 리스크 팩터에서 수익률 예측까지

7장에서 다루는 내용은 다음과 같다.

- 선형 회귀 분석의 원리와 가정은 무엇인가?
- 선형 회귀 모델 훈련과 진단은 어떻게 하는가?
- 선형 회귀 분석을 사용해 주식 수익률은 어떻게 예측하는가?
- 규제화를 사용해 예측 성과를 어떻게 개선하는가?
- 로지스틱 회귀 분석의 작동 원리는 무엇인가?
- 회귀 분석을 분류 문제로 어떻게 변환시키는가?

베이스라인 모델: 다중 선형 회귀 모델

다중 회귀 모델은 하나의 연속 결과 벡터와 어떤 유형이든 가능하지만 전처리가 필요한 입력 변수 p 사이의 선형 함수 관계로 정의된다.

데이터에서 모델 파라미터를 학습할 때 사용할 수 있는 방법에는 몇 가지가 있다. 최소자승법, 최대 우도 추정, 확률적 경사 하강법이 있다.

- 최소자승법: 모델의 예측값과 데이터 포인트에 대한 실제 출력값 사이의 차이는 잔차이다. 최소 자승 추정법은 잔차 제곱합을 최소화하기 위한 계수 벡터 β 를 선택한다.
- 최대 우도 추정: 모델 파라미터의 함수로써 주어진 두 입력 데이터에 대한 출력값 표본을 관찰하는 가능성을 계산하는 우도 함수에 의존한다. 최대 우도 추정의 목표는 주어진 입력을 취하면서 관측된 출력 표본의 확률을 최대화하는 모델 파라미터를 선택하는 것이다.
- 경사 하강법: 경사 하강법은 부드러운 함수의 정지 점을 찾는 범용 최적화 알고리즘이다. 목적 함수가 볼록인 경우 솔루션은 global optimum 이 된다.

7장 요약

장에서는 회귀 및 분류를 위한 선형 모델의 중요한 기준 사례를 이용한 첫 번째 머신러닝 모델을 소개했다. 두 과제에 대한 목적 함수의 공식화를 탐구하고 다양한 훈련 방법을 배웠으며, 추론과 예측을 위해 모델을 사용하는 방법을 배웠다. 머신러닝 기법을 활용해 리스크 관리, 새로운 알파 팩터 평가, 성과를 속성화하는 데 매우 유용한 선형 팩터 모델을 추정했다.

8. ML4T 작업 흐름: 모델에서 전략 백테스트까지

8장에서는 특정 투자 유니버스와 유용한 속성의 계산에 대한 데이터를 소싱하고 준비하는 것으로 시작한다. 이러한 특성으로부터 실행 가능한 신호를 추출하고자 머신러닝 모델을 설계하고 평가를 진행한다. 그리고 이러한 신호를 최적화된 포트폴리오로 변환하는 전략 시뮬레이션 실행과 평가로 워크플로 프로세스의 막을 내린다.

ML4T 워크플로의 궁극적인 목표는 과거 데이터에서 증거를 수집하는 것이다. 이는 후보 전략을 실제 시장에 배치하고 금융 자원을 리스크에 노출시킬지 여부를 결정하는 데 도움을 준다. 이 프로세스는 다음 포인트들을 수행할 수 있는 능력에 의존하기 때문에 이전 장들에서 개발한 기술들을 기반으로 한다.

- 다양한 데이터 세트 소스로 작업해 정보력 있는 팩터를 창출
- 트레이딩 전략의 정보력을 향상시키는 예측 신호를 생성하는 머신러닝 모델 설계
- 리스크 수익률 관점에서 최종 포트폴리오 최적화

ML 기반 전략의 백테스트 방법

백테스트는 구체적인 투자 유니버스와 기간을 사용하며, 다음 스텝들과 관련된다.

1. 시장, 기본적 및 대체 데이터 출처와 준비
2. 예측 알파 팩터와 특성 공학
3. ML 모델을 설계, 튜닝, 평가해 트레이딩 신호 생성
4. 예를 들어 규칙을 적용해 이러한 신호에 따라 거래 결정
5. 포트폴리오 맥락에서 개별 포지션의 크기 조정
6. 과거 시장 데이터를 사용해 발생한 트레이딩 시뮬레이션
7. 최종 포지션이 어떻게 수행됐는지의 평가

백테스트 엔진 작동법

간단히 말해 백테스트 엔진은 과거 가격에 대해 반복 실행하면서 현재 값을 알고리즘에 전달하며, 그 대가로 주문을 받고 결과 포지션들과 이들의 값을 추적한다. 벡터화 접근법과 이벤트 기반 접근법 간의 차이는 실제 트레이딩 환경의 충실한 복제가 어떻게 상당한 복잡도를 가중시키는지 보여준다.

[벡터화 백테스트 & 이벤트 기반 백테스트]

벡터화 백테스트는 전략을 평가하는 가장 기본적인 방법이다. 그것은 단순히 목표 포지션 크기를 나타내는 신호 벡터에 투자 기간의 수익률 벡터를 곱해 기간 성과를 계산한다.

이 책에서는 예측들을 매우 간단한 전략을 위한 신호로 바꾼다. 어떤 거래일에도 10 개의 가장 높은 긍정적인 예측에 대해 롱을 취하고, 가장 낮은 10개의 부정적인 예측에 대해 숏을 취한다.