해당 EMP는 다음과 같이 구성된다.

기본적으로 Target Value는 7영업일 뒤의 해당 ETF 수치이다. 단기 리밸런싱 위주로 운용하며 유니버스는 Blackrock에서 상장한 iShares 시리즈 중 거래비용, Aum, 거래량을 고려하여 선정하였다.

|  |  |  |  |
| --- | --- | --- | --- |
| ETF Name | Aum | Expense Ratio | Incept. Date |
| iShares Core MSCI EAFE ETF | 73,307M | 0.07% | 2012.10.18 |
| iShares Core MSCI Emerging Markets ETF | 60,945M | 0.14% | 2012.10.18 |
| iShares Core S&P 500 ETF | 202,175M | 0.04% | 2000.5.15 |

해당 ETF 상장 이전 시기의 경우 같은 Index를 추종하는 ETF 데이터로 대체하였다. 이후 ETF의 수치를 예측하여 기대수익률을 스코어로 투자한다.

예측은 여러 예측 알고리즘의 블렌딩으로 이루어진다. 알고리즘은 Ridge, Lasso, ElasticNet, Gradient Boosting, Multi-Layer Perceptron 으로 나뉘며 Gradient Boosting의 경우 sckit-learn의 GradientBoostingRegressor 함수와 XGBoost를 모두 사용하였다. 각각의 알고리즘으로 예측값을 도출한 후, Train-set 정확도를 Weight로 가중평균을 도출한다.

Feature의 경우 다음 데이터를 사용한다.

WTI, Gold, EAFE, EM, SP500, Copper, Nickel, Aluminium, Zinc, Canada Sp, FTSE, Shanghai, Nikkei225, Hangseng, BVSP, Kospi, SP500 future, BVSP Future, FTSE Future, USD, EUR, CNY, JPY, GBP, Canada60 Future

해당 데이터에 대해 결측치는 이전 일자의 데이터로 채운다. 이후 5일, 15일, 60일, 150일 이동평균을 계산하여 Feature로 추가한다. 총 칼럼 길이 2622 row가 생성되었다.

모델에 입력할 데이터는 Sliding Window 방식으로 구성한다. 각 Window의 길이는 Train-set은 150영업일, Test-set은 7영업일로 두었다. 즉, x\_train\_set의 경우 150 row를 갖는 331개의 매트릭스로 구성된 리스트, x\_test\_set의 경우 7 row를 갖는 331개의 매트릭스로 구성된 리스트가 된다. 이후 각 매트릭스 별로 데이터 칼럼 내에서 분산을 동일하게 만들기 위해 0~1 사이의 값으로 변환한다. 매트릭스가 구분되어있기 때문에 미래참조는 발생하지 않는다. Train-set 길이의 경우 Tuning에 시간이 대단히 많이 소요되어 적정값을 구하지는 못하였으나 길이가 지나치게 짧을 경우 큰 오차가 발생하였고, 길이가 지나치게 길 경우 시기적 특성을 대변하지 못하는 결과가 도출되었다. 따라서 150영업일을 기준으로 하였다. Test-set의 경우 투자결정에 영향을 주는 바로 이어지는 1영업일에 대한 데이터만 있으면 되지만 7영업일 이후의 수치를 예측하는 것이기 때문에 해당 기간에 대한 설명력을 가져야한다고 생각하여 Test-set 길이를 지정하였다.

Loss function의 경우 (mean(Train-set-score) + mean(Test-set-score)) / 2 로 정의하였다. Train-set의 경우 150개 row를 갖기 때문에 150개의 yhat값에 대한 오차가 1개의 값으로 변환되어 반영되고, Test-set의 경우 7개의 row를 갖기 때문에 7개의 yhat값에 대한 오차가 1개의 값으로 변환되어 반영된다. 따라서 Test-set이 loss-fuction에 더 큰 영향을 주도록 구성하였다.

HyperParameter Tuning의 경우, 예측력에 영향을 주는 Parameter를 선별하여 HyperParameter Space를 정의하였다. 따라서 해당 Space내에서 최적값을 도출하도록 설정하였다. 이는 각 Window에 따라 각각 다른 수치가 기록됨을 뜻하고, 각각의 시점을 잘 설명하는 모델이 다를 것이라는 생각으로 블렌딩을 실행하였다.

본 알고리즘의 실효성을 체크하는 것이 주 목적이므로 포트폴리오는 간단하게 비중조절이 아닌 단일자산 투자로 구성한다. 예를들어 2009년 5월 20일 기준 MSCI EM의 스코어가 가장 높다면 100% MSCI EM에 투자한다고 가정한다. 앞서 밝힌 ETF의 거래비용을 매 거래마다 적용하였으며 리밸런싱은 7영업일마다 실시하였다.