

Nanodegree Engenheiro de Machine Learning

Proposta de projeto final

Thiago Henrique Martins de Souza

18 de março de 2019

Proposta

Histórico do assunto

Empresas de investimento, hedge funds e até indivíduos têm usado modelos financeiros para entender melhor o comportamento do mercado e fazer investimentos e negócios lucrativos. Uma riqueza de informações está disponível na forma de preços de ações históricos e dados de desempenho da empresa, adequados para algoritmos de aprendizado de máquina a serem processados.

Descrição do problema

Realizar a previsão do valor do fechamento ajustado para ações de um determinado período, utilizando dados históricos como entrada. Um script simples deverá ser capaz de prever o valor de ações e compará-lo com os valores reais, mostrar isso através de um gráfico e cálculo da média total.

Conjunto de dados e entradas

Os dados serão extraídos da fonte Yahoo Finanças, ações da Petrobras PETR4.SA no período de 01/01/2013 até 31/01/2013. Arquivos .csv utilizados estarão anexo ao projeto enviado.

Descrição da solução

Para este projeto, criarei um preditor de preço de ações (script simples) que utiliza dados de negociação diários em um determinado período como entrada e gera estimativas projetadas. Observe que as entradas conterão múltiplas métricas, como preço de abertura (Aberto), preço mais alto das

ações negociadas em (Alta), quantas ações foram negociadas (Volume) e preço de fechamento ajustado para desdobramentos e dividendos. Onde será previsto apenas o preço do fechamento ajustado. O preditor deverá ser capaz de prever os valores o mais próximo possível do valor real no período. Para isso irei utilizar uma rede neural recorrente, LSTM do Keras com 4 (quatro) camadas e 100 (cem) épocas.

Métricas de avaliação

Para avaliar o desempenho do projeto, podemos verificar o valor da média geral entre valor real e valor previsto e será gerado um gráfico com os valores reais e os valores previstos. A intenção será acertar o maior número de vezes o valor da ação e/ou tentar ficar o mais próximo possível do valor real da ação.

Design do projeto

O projeto seguirá um fluxo de trabalho simples mas eficiente, primeiramente os dados serão normalizados utilizando Sklearn MinMaxScaler, depois serão convertidos em Numpy Array, depois os dados normalizados serão processados pela rede neural, os valores previstos serão extraídos e comparados com os valores reais. Comparando média geral e visualizando o gráfico Valor Real X Previsões.

Recursos

Keras LSTM: <https://keras.io/layers/recurrent/>

Python for Finance:

<https://www.learndatasci.com/tutorials/python-finance-part-yahoo-finance-api-pandas-matplotlib/>