香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

MATH 5470 Final Project
(Re-)Imag(in)ing Price Trends

Hu, Mingyun
Ma, Wanteng
Zhang, Jiaxin

Youtube link: https://youtu.be/9YdI_NMriSw

Introduction

- Literature has proven the feasibility of using **past prices** to predict **future returns**.

- Jiang et al. use the convolutional neural network (CNN) to learn about the behavior of price trend based on pattern recognition.

Tesla OHLC chart with daily trading volume

Data Preprocessing

"Imaging" Market Data:

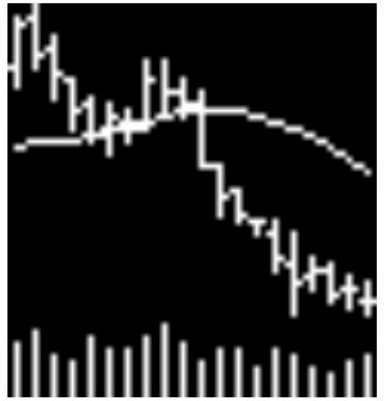
The key idea is to transfer 1-D market data to 2-D OHLC images.

High price, low price, open price and close are shown in each day.

We investigate the 20-day return with image data which are of size 64x60 and each one features moving average line (MA) and volume bar (VB).

The labels are the returns of the next time horizon with 1 indicating positive return and 0 otherwise.

OHLC and image sample



Dow Jones Industrial Average (^DJI)
DJI - DJI Real-time price. Currency in USD

Add to watchlist

32,977.21 -939.18 (-2.77%)
At close: 04:59PM EDT

Summary | Chart | Conversations | **Historical data** | Options | Components

Time period: 30 Apr 2021 - 30 Apr 2022 | Show: Historical prices | Frequency: Daily | Apply

Currency in USD

Date	Open	High	Low	Close*	Adj. close**	Volume
29 Apr 2022	33,787.01	33,919.59	32,913.15	32,977.21	32,977.21	501,560,000
28 Apr 2022	33,425.96	34,054.79	33,248.46	33,916.39	33,916.39	440,380,000
27 Apr 2022	33,450.92	33,697.18	33,108.89	33,301.93	33,301.93	447,230,000
26 Apr 2022	33,907.49	33,909.51	33,230.95	33,240.18	33,240.18	400,020,000
25 Apr 2022	33,731.65	34,106.01	33,323.37	34,049.46	34,049.46	416,900,000

Digital market data: OHLC+volume

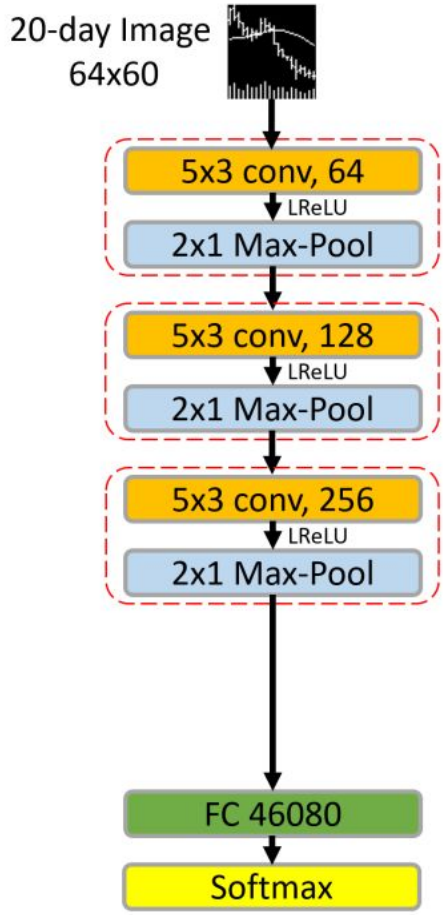
CNN Baseline Model

> 3 layers
    each layer contains
        5 x 3 convolution
        batch normalization
        leaky ReLU activation
        2 x 1 max pooling
    output an 1D vector

> dropout

> a fully connected layer
    output the probabilities of all classes

> softmax



Sensitivity Analysis of Model Structure

Model		Validation Loss	Validation Accuracy	Test Accuracy
Baseline		0.693	0.546	0.528
Dropout (0.50)	0.00	**0.687**	0.545	0.518
	0.25	0.688	0.547	0.533
	0.75	0.694	**0.548**	0.531
Xavier (yes)	no	0.690	0.544	0.518
Filter Size (5×3)	(3×3)	0.691	0.547	0.498
	(7×3)	0.690	0.543	0.511
Filters (64)	32	0.714	0.545	0.548
	128	0.689	0.542	0.542
Layers (3)	2	0.754	0.526	0.533
BN (yes)	no	0.697	0.544	0.445
Activation (LReLU)	ReLU	0.697	0.540	**0.551**
Max-pool Size (2×1)	(2×2)	0.702	0.538	0.525

Sensitivity to Model Structure

The highest validation accuracy appears when the dropout probability equals to 0.75, which indicates the robustness of CNN model.

Decreasing the number of layers meaningfully reduces the model performance.

Sensitivity Analysis of Model Structure

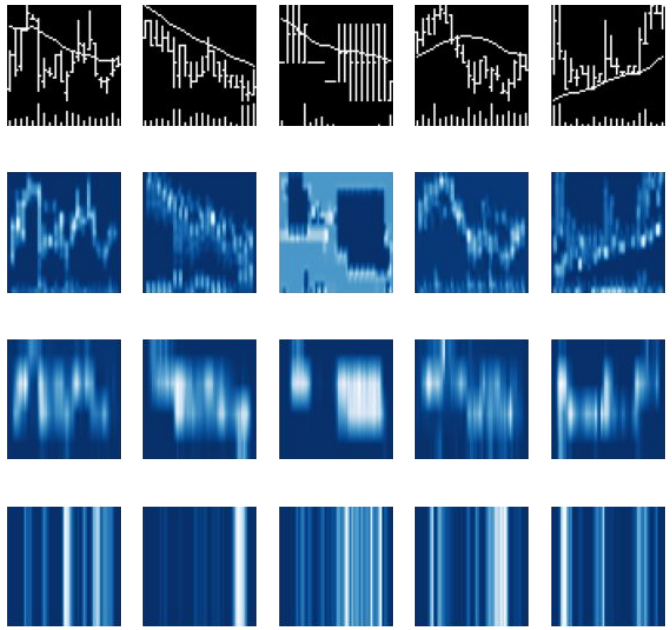
Model		Validation Loss	Validation Accuracy	Test Accuracy
Baseline		0.693	0.546	0.528
Dropout (0.50)	0.00	**0.687**	0.545	0.518
	0.25	0.688	0.547	0.533
	0.75	0.694	**0.548**	0.531
Xavier (yes)	no	0.690	0.544	0.518
Filter Size (5×3)	(3×3)	0.691	0.547	0.498
	(7×3)	0.690	0.543	0.511
Filters (64)	32	0.714	0.545	0.548
	128	0.689	0.542	0.542
Layers (3)	2	0.754	0.526	0.533
BN (yes)	no	0.697	0.544	0.445
Activation (LReLU)	ReLU	0.697	0.540	**0.551**
Max-pool Size (2×1)	(2×2)	0.702	0.538	0.525

Sensitivity to Model Structure

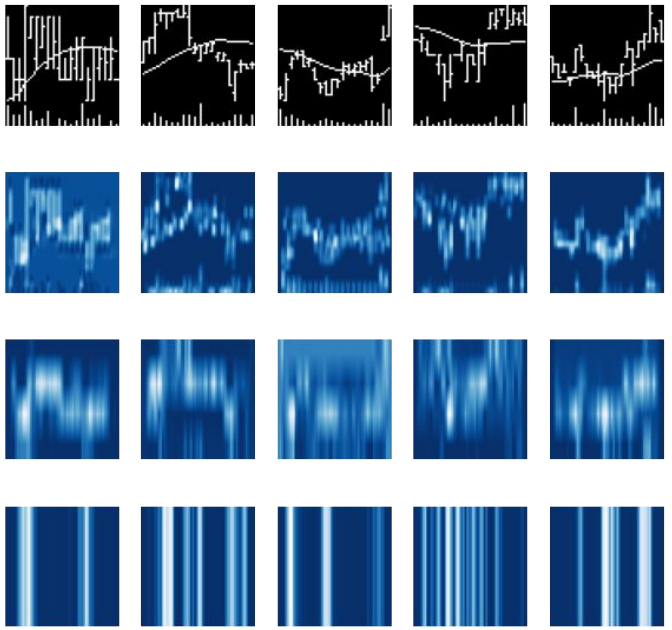
The model performance is fairly sensitive to the use of batch normalization.

The optimal structure is when we transition from leaky ReLU to ReLU.

CNN Model Interpretation



Samples classified as 0



Samples classified as 1

We try to visualize and interpret the model by leveraging grad-CAM.

(1) the first layer can capture the inflection points of the trends;

(2) Open and close prices are of interest by model;

(3) The volume information can be reflected in the second and last layers. The trends are highlighted by the volume bar of days.

CNN Regression Model

Loss Function: Mean Square Error

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Accuracy: R^2 statistic

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$
$$where\ TSS = \sum(y_i - \bar{y})^2, RSS = \sum(y_i - \hat{y}_i)^2$$

Regression Model		Validation Loss	Test Loss
Baseline		0.05248	0.02665
Dropout(0.5)	0.00	**0.05231**	0.02669
	0.25	0.05262	0.02668
	0.75	0.05260	0.02650
BN(yes)	no	0.05239	0.02663
Activation(LReLU)	ReLU	0.05242	**0.02649**

Thank You