

Stock Market Prediction via Twitter Sentiment Analysis

Bobby Swingler, Isai Mercado Oliveros, Isaac Kurth

Idea

- Predict Apple's Stock price based on what people say about Apple and its competitors in twitter.



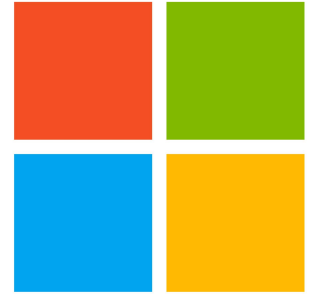
Approach

- Define what companies, we think, relate to apple so that their opinions are relevant for predicting Apple's stock



Google

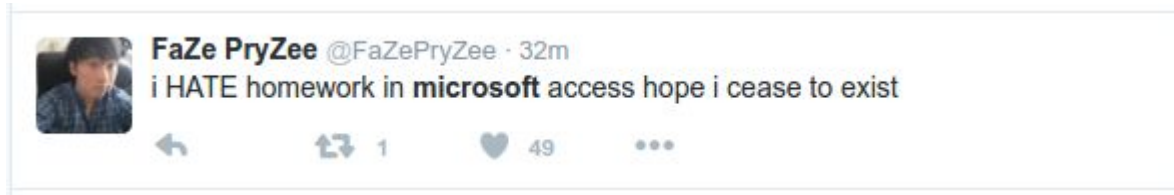
SAMSUNG



SONY

Get Tweets

- Gather and sort these companies tweets by day



Tweet Cleaning

- Remove URLs and untokenizable characters
- Remove hashtags
- Remove user's "@" symbol
- Relevant to Apple company, not the fruit



Get Sentiment Score

- Score tweets with Sentiment Analyzer



O'Reilly Media @OReillyMedia · 3h


How does **Google** make sure it (almost) never goes down? 3 words: Site Reliability Engineering. Read the Wired article wired.com/2016/04/google...



- 0 -> "Very Negative"
- 1 -> "Negative"
- 2 -> "Neutral"
- 3 -> "Positive"
- 4 -> "Very Positive"

Calculate mean per day

- Calculate mean for all companies per day and enter entry in data set

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$


Date	Day	Apple	Microsoft
3/1/2016	Tue	1.373250389	1.419354839

Calculate Target

- Calculate stock fluctuation by subtracting the market close price from the market open price for the given day



-

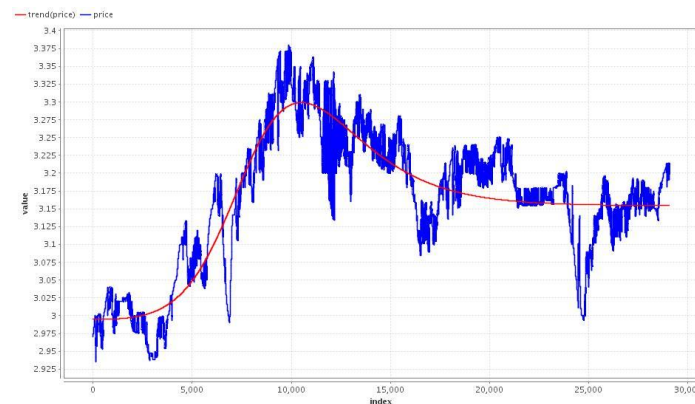
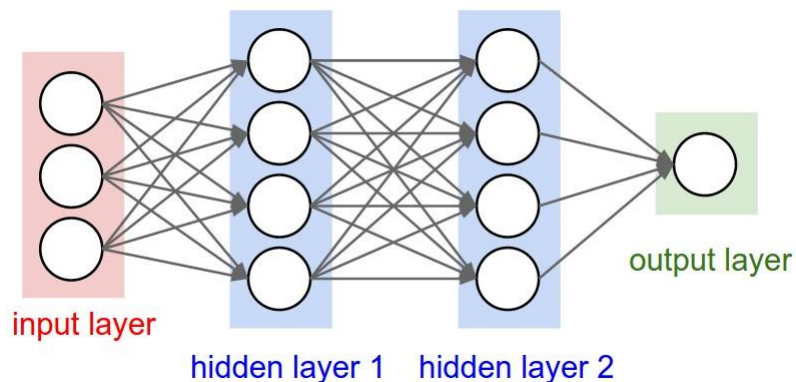


=



Model 1 - Neural Network

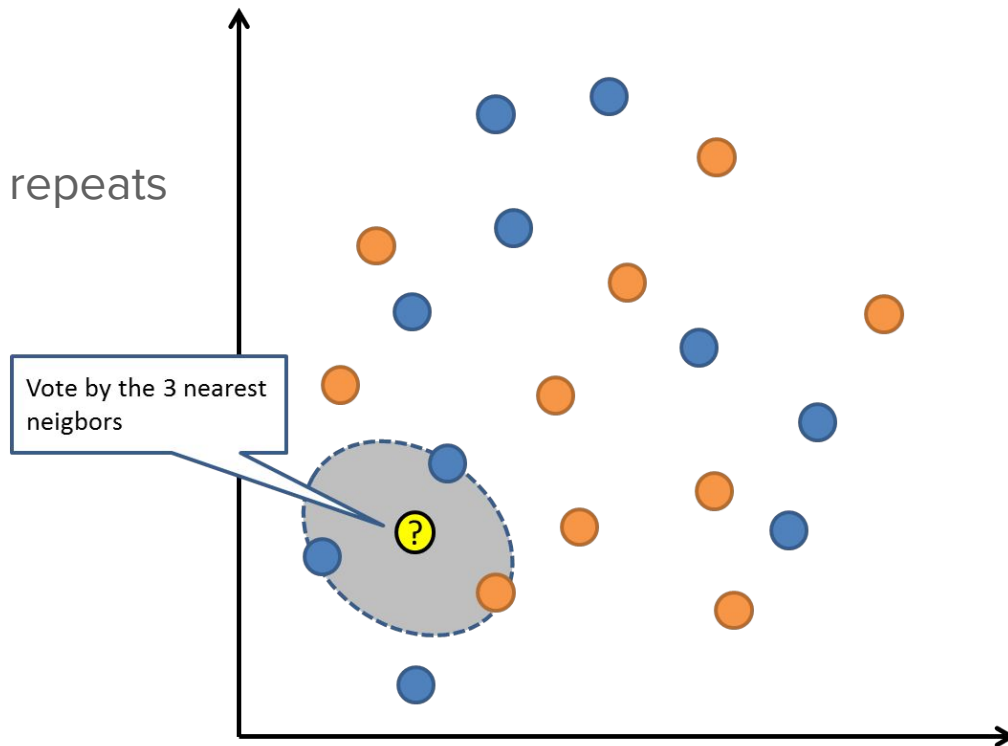
- Assumption: Relationships between opinions and stocks are very complicated



10 Fold Cross Validation Test Set RMSE = 0.4 Cents

Model 2 - KNN

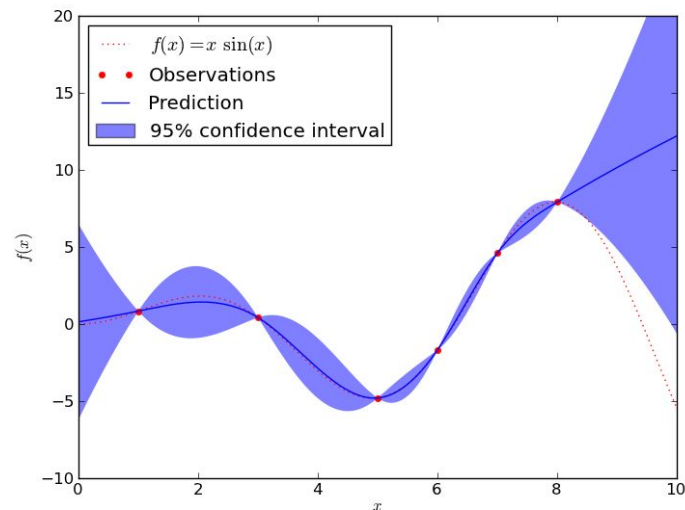
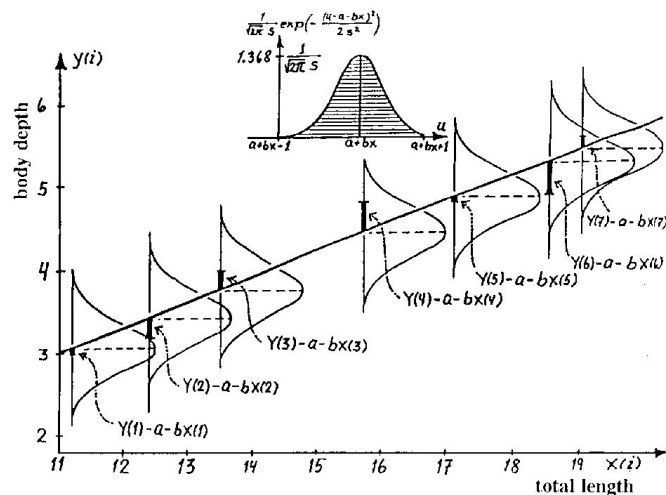
- Assumption: History repeats



Test Set RMSE = 0.78 Cents

Model 3 - Gaussian Regression

- Assumption: there is a distribution of sentiments. Not just a MEAN



Test Set RMSE = 0.9 Cents

Model 4 - Linear

Assumption - Maybe we just need overall monthly trend



Test Set RMSE = 0.94 Cents

3 Days for Testing

- After our month of training, we tested for the next three days

Real Life Test(3 Days after our training month)						
Date	Day	Target	Neuron Net	KNN	Gaussian	Linear
4/4/2016	Mon	.7	.6	.0	.6	.2
4/5/2016	Tue	-.6	-1.1	1.5	1.0	2.5
4/6/2016	Wed	1.2	4.8	.6	.6	1.1

- N Net direction was always correct, but magnitude got worse as days passed

3 Days for Testing

- After our month of training, we tested for the next three days

Real Life Test(3 Days after our training month)						
Date	Day	Target	Neuron Net	KNN	Gaussian	Linear
4/4/2016	Mon	.7	.6	.0	.6	.2
4/5/2016	Tue	-.6	-1.1	1.5	1.0	2.5
4/6/2016	Wed	1.2	4.8	.6	.6	1.1

- N Net direction was always correct, but magnitude got worse as days passed
- KNN did not have enough neighbors to compare (training set was 30 rows)

3 Days for Testing

- After our month of training, we tested for the next three days

Real Life Test(3 Days after our training month)						
Date	Day	Target	Neuron Net	KNN	Gaussian	Linear
4/4/2016	Mon	.7	.6	.0	.6	.2
4/5/2016	Tue	-.6	-1.1	1.5	1.0	2.5
4/6/2016	Wed	1.2	4.8	.6	.6	1.1

- N Net direction was always correct, but magnitude got worse as days passed
- KNN did not have enough neighbors to compare (training set was 30 rows)
- Gaussian did fine first and third days, but not trustable because of second day

3 Days for Testing

- After our month of training, we tested for the next three days

Real Life Test(3 Days after our training month)						
Date	Day	Target	Neuron Net	KNN	Gaussian	Linear
4/4/2016	Mon	.7	.6	.0	.6	.2
4/5/2016	Tue	-.6	-1.1	1.5	1.0	2.5
4/6/2016	Wed	1.2	4.8	.6	.6	1.1

- N Net direction was always correct, but magnitude got worse as days passed
- KNN did not have enough neighbors to compare (training set was 30 rows)
- Gaussian did fine first and third days, but not trustable because of second day
- Linear Regression does not transmit too much information

Future Work

- Twitter Archive
- Model Optimizations
- Further Sentiment Analysis and testing