Making It Rain with Machine Learning

James Kazmierczak and Neil Shah

Mentors: Chae Clark and David Darmon

- Motivation
- Financial Time Series
- Best Practices for Learning from Data
- Ensemble Methods
- Baseline Model
- James’s Model
- Neil’s Model
- Results

Motivation

Create a model that predicts the stock market greater than 50%

Learn about machine learning and statistical learning

Make money through prediction

Introduction into financial mathematics field

Learn and apply ensemble methods

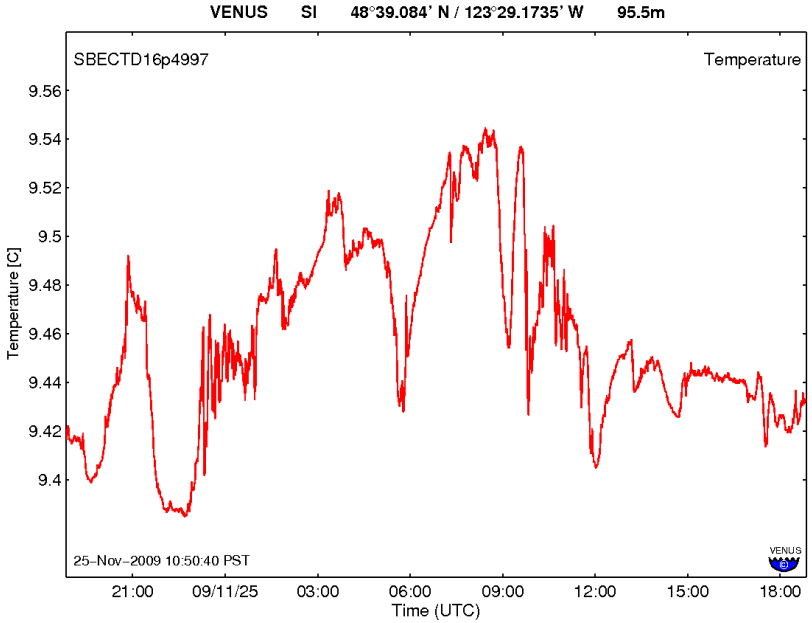
Time Series Background

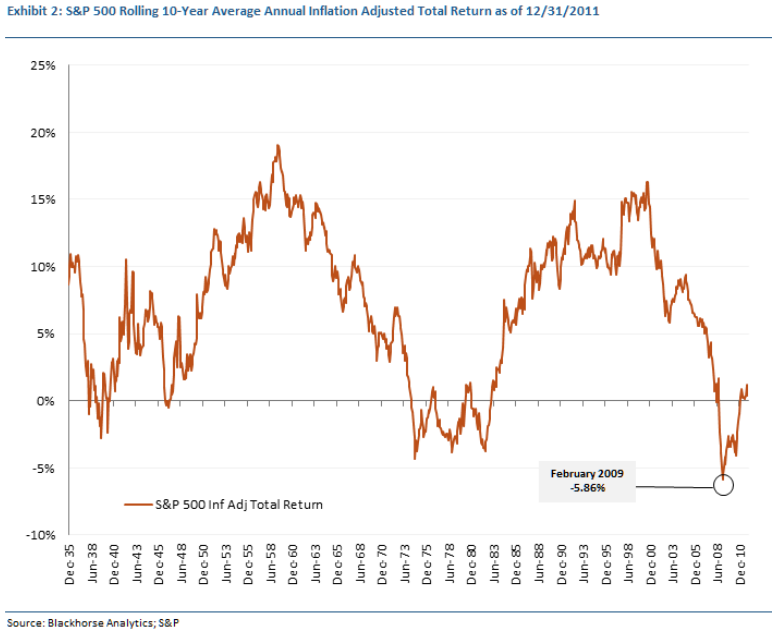
Using knowledge from the past to predict the future

Deterministic - The future is a function of the past. These models are imperfect because the data is incomplete

Stochastic - Stochastic model gives a probability distribution of the future based on the past.

Time Series Background (cont.)





Learning Background

Basic definition: predicting the future based on the past

Accomplished by assuming a form for the model and learning its parameters so that model can be generalized and used to predict the future

Learning Background (cont.)

Features and Outcomes

Feature Space: X

Outcome Space: Y

f maps from the features to the outcomes:

$$f: X \rightarrow Y$$

Learning Background (cont.)

Division of the data:

- Training set: data on which our model is expected to learn
- Tuning set: making small changes to the model to improve prediction
- Testing set: data on which our model is tested for accuracy

Example: Apples and Oranges

Ensemble Methods

Crowd vs Expert

Only works if unbiased uncorrelated

Application to models

Instead of using just one model to predict, we use multiple uncorrelated models so when one fails they do not all fail. Then the average of all the models will be able to give us the answer

Stock Prediction Background

Price of a stock: value investors are willing to pay for a share of the stock

Stock return: the gain or loss incurred by an investor over a period of time

Volatility: the amount of uncertainty or risk about the size of changes in a stock's value

- Generally, higher volatility=higher risk

Stock Prediction Background (cont.)

Regime switching: allows forecasting model to predict stock price to switch between K regimes, with each regime characterized by different model parameters.

Moving average: trend following lagging indicator that helps smooth out “noise” from random stock price fluctuations

Binary prediction: use of two feature vectors in creation of learning model

Autoregression: model where output variable depends linearly on its own previous value

Random Walk Model

Assume the market is a random walk:

Related to the Efficient Market Hypothesis

Then stock returns act like a biased coin.

Always predict the biased value.

Up-Down Model

Divide data into training, tuning, and testing sets.

Use different lengths for looking into the past.

Used tuning set to determine the best length

Used the best length to predict the testing set

Umbrella Model — James

Take the weekly average of the stock prices Take the differences and range from 0 to 5

Based on past performances predict the future

Run looking at past 2 and 3 weeks

This will predict the number of ups in the upcoming week.

Then using this as a limit. Average other indicators like daily up-down, did the daily cross the moving average, and major news event in the past 10 days.

Predict Up or down until you reach the predicted ups.

Model 2 — Neil

International markets (weight: 0.25)

- Look at Frankfurt's DAX 30, London's FTSE 100, Japan's Nikkei, Shanghai Composite, and S&P500

5 day average (weight: 0.5)

- Look at 5 day closing average and compare to previous day's close

Volatility (0.25)

- Look at how stock has performed historically based on positive/negative volatility

Results

Stock Name	P(Up)	Random Walk Acc.	Up-Down Acc.	Mentor Acc.
Exxon	0.513834	0.486166	**0.509960**	0.4683
Ford	0.494071	0.482213	**0.502041**	0.4921
McDonald's	0.569170	**0.553360**	0.547325	0.5516
YUM	0.545455	0.498024	**0.518519**	0.4960