Machine Learning for Trading
Financial Investing

Dealing with Data

How Data is aggregated

• Many trades
• Different Exchanges
• Tick
  – Finest resolution of data (minimum size)
    • Price (1c)
    • Upward, downward movement (change in price)
  – Cell Match
    • Buy and Sell Match

How Data is aggregated

• Many trades
• Different Exchanges
• Tick
  – Finest resolution of data (minimum size)
    • Price (1cent)
    • Upward, downward movement (change in price)
  – Cell Match
    • Buy and Sell Match

Data

• Historical
• Price
• Volume
• Lots of Data
• Consolidate Data
  – Time Epoch
    • Minute by Minute
    • Hour by Hour
    • Daily
  – Open, Close
  – High, Low
  – Volume
  – Combine different exchanges

Price anomaly

• Example 1:
  – Blue Line:
  – What are the price drop here?

• Example 2:
  – And here, what are these price drops?

Which makes most sense?

- CEO Quit
- Dividend Cut
- Stock Split
Price anomaly

- CEO Quit
- Dividend Cut
- Stock Split

- Next day 4 shares of 75, $4\times75=300.
- Next day 2 shares of 125, $2\times125=250.

Why Split?

- Less Liquid, less volume
- Options on stocks are usually traded with regard to 100 shares.
- Finely Tuned portfolio (harder with high priced stocks).

Why do Stock Splits?

- Price becomes to high: less liquid, less volume.
- Buy in groups of 100 (typical)
  - 50,000 for $500/stocks.
- Exceptions/Why make an Exception?
  - Berkshire Hathaway BRK.A. (doesn't split it shares).
    - Warren Buffet's holding company.
    - $110,000 (Volume: 450 shares per day)
    - Keeps the price high to deter short time traders creating excess volume.
  - Seabord SEB, $2,660
  - NVR $703
  - GOOG -- $618 no splits. Started at $100 in 2004

How to deal with stock split data

- Not short stocks when company value has not change:
  - Adjusted Close.
  - Current Day:
    - Adjusted close = close
    - Back in history need to adjust it.

https://www.forbes.com/sites/investor/2011/07/25/bershire-seaboard-google-priceline/7f852b8d8d4ab

2017 prices
• How to deal with stock split data
  • Not short stocks when company value has not change:
    – Adjusted Close.
    – Current Day:
      • Adjusted close = close
      • Back in history need to adjust it.

• Day before and day off?

• Answer:
  – $101 day before
  – $100 day after.

• Dividends,
  – Annually
  – Quarterly

• Example
  – $100.00
  – $1.00 dividend announced (1%)
  – Stock value?
    • Assume consensus of stock vale is $100.00 (may not be the price).

Dividends

Here we have:
$1 + $100 share
1) What is the value of the stock the day before dividend payout?
2) What is the value on they day of payout?

Stock value?

$100.00 (may not be the price)
Adjusting for Dividends

- Similar to stock splits.
- Red line is Adjusted close.

Survivor Bias

- SP 500
  - Membership Change.
- Simulate todays list in the past using historical data
  - Current list survived in the past.
- Survivor Bias-
  - Inflation of funds that remain when poor performers are not part of the equation.
- Need survivor bias free data (costs money).
  - To more accurately measure strategies working on historical data.
  - Need to know historical lists of companies.