What to do before machine learning?

Ernie Chan, Ph.D.
QTS Capital Management, LLC.

ML and its perils

- Any smart high schooler knows how to run a random forest or deep neural net on financial data.
- Your value-add is what to do before and after.
- I will focus on the “before”.

3 Steps

- What are you trying to predict?
- Benchmark/Baseline strategies.
- Data/Features engineering.

What are you trying to predict?

- Use ML to predict things that are not subject to “reflexivity”. (*George Soros*)
  - Returns is reflexive.
  - Weather in the midwest is not.
  - Gasoline consumption is not.
  - Realized volatility is not. (Question: why not?)
  - Recession is not. (Question: are we sure?)

What are you trying to predict?

- Even if you want to predict returns
  - What time horizon?
  - Do you want to predict just sign, or magnitude as well?
  - Is long or short time horizons easier to predict?

Benchmark/baseline Strategies

- E.g. the baseline sign-of-return predictive accuracy in financial is not usually 50%.
  - Returns series is usually either trending or mean-reverting.
  - I.e. serially correlated or anti-correlated.
  - Baseline strategy is a simple momentum or mean-reverting strategy.

Benchmark/baseline Strategies

- Next: linear or logistic regression, with single predictor.
- Next: classical time series models such as ARIMA.
- Next: linear or logistic regression, with multiple predictors.
- Next: linear or logistic regression, with L1 regularization. (Question: what is L1?)
- (Best predictive performance usually a combination of shallow and deep learning. – *François Chollet*)

Benchmark/baseline Strategies

- Next: “manual” quant strategies..
- ML can often be used to improve on non-ML quant trades via “meta-labelling” (Lopez de Prado, 2018).

Data Engineering

- First step in any ML models: check data integrity.
  - Noisy/wrong data?
  - Missing data?
  - Retroactive revisions?
    - E.g. Often earnings announcement dates are not Point-In-Time!
    - (Companies revise expected announcement dates up till day of expected announcement.)
  - Look ahead bias in features?
  - Are features stationary?
    - E.g. Cannot use price as features.
    - Need “fractional differentiation”.
  - Are features synchronous?
    - E.g. Cannot simply combine daily closing prices of stocks with futures and options.

Questions

1. What is the problem with using, say, I/B/E/S earnings announcement dates for backtesting an event-driven model?
2. Is return a stationary variable? Is it suitable as feature?
3. Can you use closing value VIX index on day t as feature to trade SPY at stock market close of same day?

Answers

1. I/B/E/S historical earnings data only tells you the actual announcement date, not the expected one.
   - We often have to enter trade based on expected date.
2. Returns are stationary – suitable as feature.
3. VIX index closing value is obtained at 16:15 ET. SPY closes at 16:00 ET.

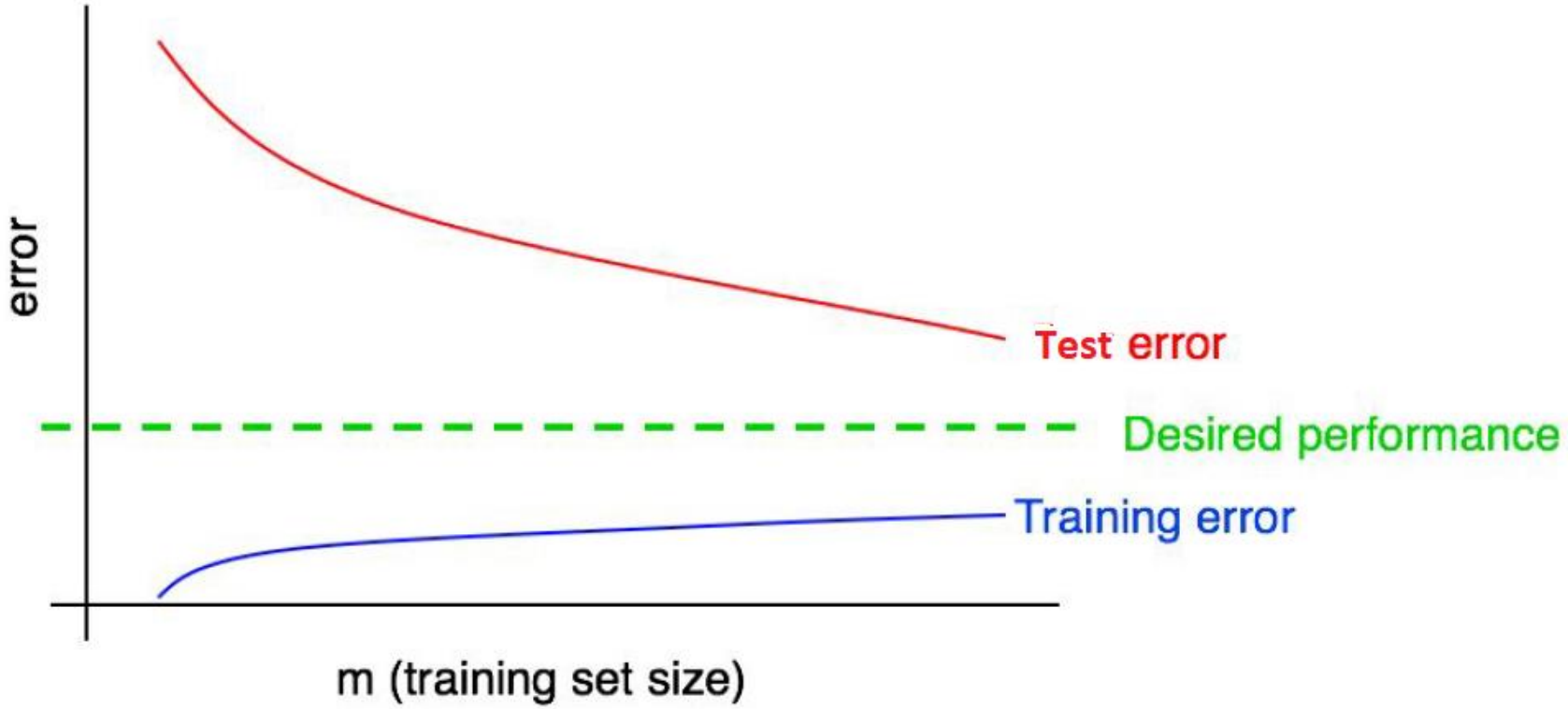
Data Engineering

- Non-price data: especially tricky to ascertain quality.
  - Best to extract features from raw data (e.g. news from Thomson Reuters) instead of relying on 3rd party sentiment models.
- More data reduces “variance”
  - But how much is enough?
  - Plot “learning curve”

Questions

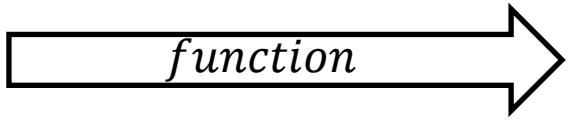
1. Does more data also reduce “bias”?
2. What is a “learning curve”?

Learning curve



Source: Andrew Ng, “Machine Learning Yearning”

Features Engineering

Raw data  Features

Question: What possible $function$ can you think of?

Examples of *function*

1. Lagged values, lagged differences.
2. Technical indicators.
3. Signal processing (e.g. Fourier transform).
4. Time series models coefficients.
5. Parametric distributions parameters.
6. Products and $function$ of $function$.

Features Engineering

- This is not as important for deep learning. (Why?)
- Quite important for shallow learning and random forests.

Features selection

- Too many features reduce bias but increase variance.
  - Lead to overfitting.
- Many methods of features selection.
- Question: which features selection method we have already discussed?
  - What other features selection method are there?

Answer

- We already discussed l_1 regularization.
  - That is mainly applicable to linear models.

Features selection

- Stepwise regression
  - Add features one at a time to maximize BIC.
  - Drop features one at a time to maximize BIC.
- MDA: mean decrease accuracy
  - Randomizes (permutes) the values of one feature at a time, and note how much this decreases out-of-sample predictive accuracy.
  - Larger the decrease in accuracy -> higher feature importance.

Finally, …

- You are now ready to apply your random forest, SVM, or neural network to your features.

Thank you for joining us!

www.epchan.com
Blog: epchan.blogspot.com
Twitter: @chanep