



MACHINE LEARNING ALGORITHMS FOR TRADING

Lesson 1: How Machine Learning is used at a hedge fund

- introduce problem early
- Overview of use and backtesting
 - Out of sample
 - Roll forward cross validation
- Methods
 - Linear regression
 - KNN regression
 - Decision trees Random Forest regression (considering to drop)
- Quiz: which algorithm makes most sense here?
- Supervised ML (intent is that the treatment here is light)
 - Use: Regression
 - Use: Classification
 - Model type: Parametric
 - Model type: Instance-based
- Quiz: What's the next point?
- Problems with regression for finance
 - Hint at reinforcement learning
- Introduce the problem we will focus on in the rest of the class, namely:
 - Example data, will learn on over a particular year (2012)
 - Will test on over the next two years (2013 2014)
 - It will be "easy" data that has obvious patterns
 - You will create trades.txt and run them through your backtester

Lesson 2: Regression

[note: need to create fake stock data that has embedded patterns]

- Overview of how it fits into overall trading process
- Definition of the problem 1
 - Black box diagram
 - training: X_{train} , Y_{train}
 - using: Query with X
- Definition of the problem 2: APIs
 - constructor
 - `addEvidence(X,Y)`
 - `query(X)`
- How to implement linear regression

Lesson 3: Assessing a learning algorithm

- Now that we have two, (linreg & KNN), let's compare them
 - Pros and cons of LinReg versus KNN
 - LinReg can extrapolate
 - Kernel
 - Piecewise
 - ease of adding new data
- Cross validation,
- roll forward cross validation
 - Use all data versus most recent data
 - Online learning
- How long to take to learn versus query
- Batch versus online
- RMS error
- Scatterplot predict vs actual
- `Corrcoef`
- Overfitting

Lesson 4: Ensemble learners, bagging and boosting

Discuss ensembles, show that ensemble learners can be ensembles of different algorithms. Netflix Prize.

Mention that this could mean different algorithms.

Bagging is an easy way to do this.

Boosting

perhaps include decision trees.

Lesson 5: Reinforcement Learning

- Classic view of the problem (from Kaelbling, Littman, Moore)
- Model-based
- Model-free

Lesson 6: Q-Learning

Lesson 7: Dyna