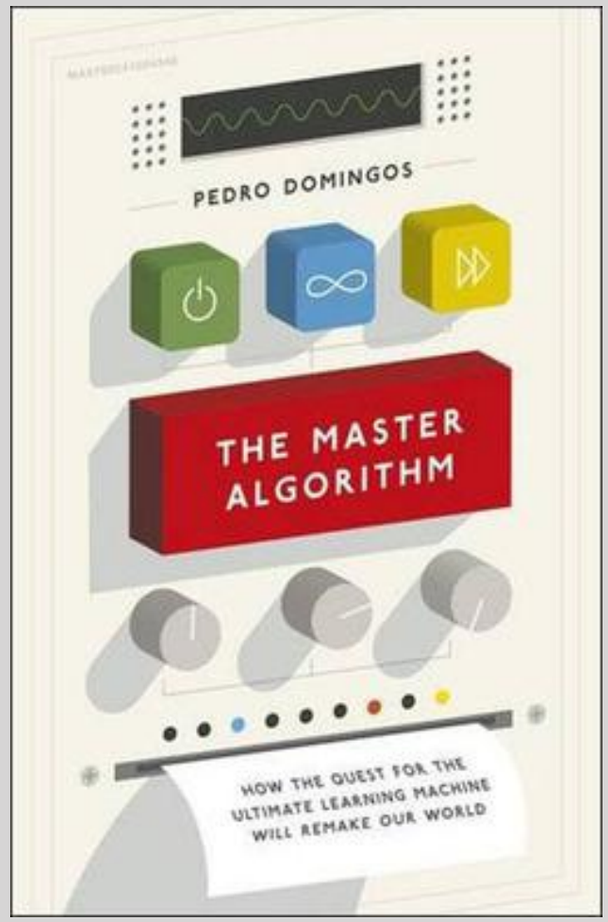
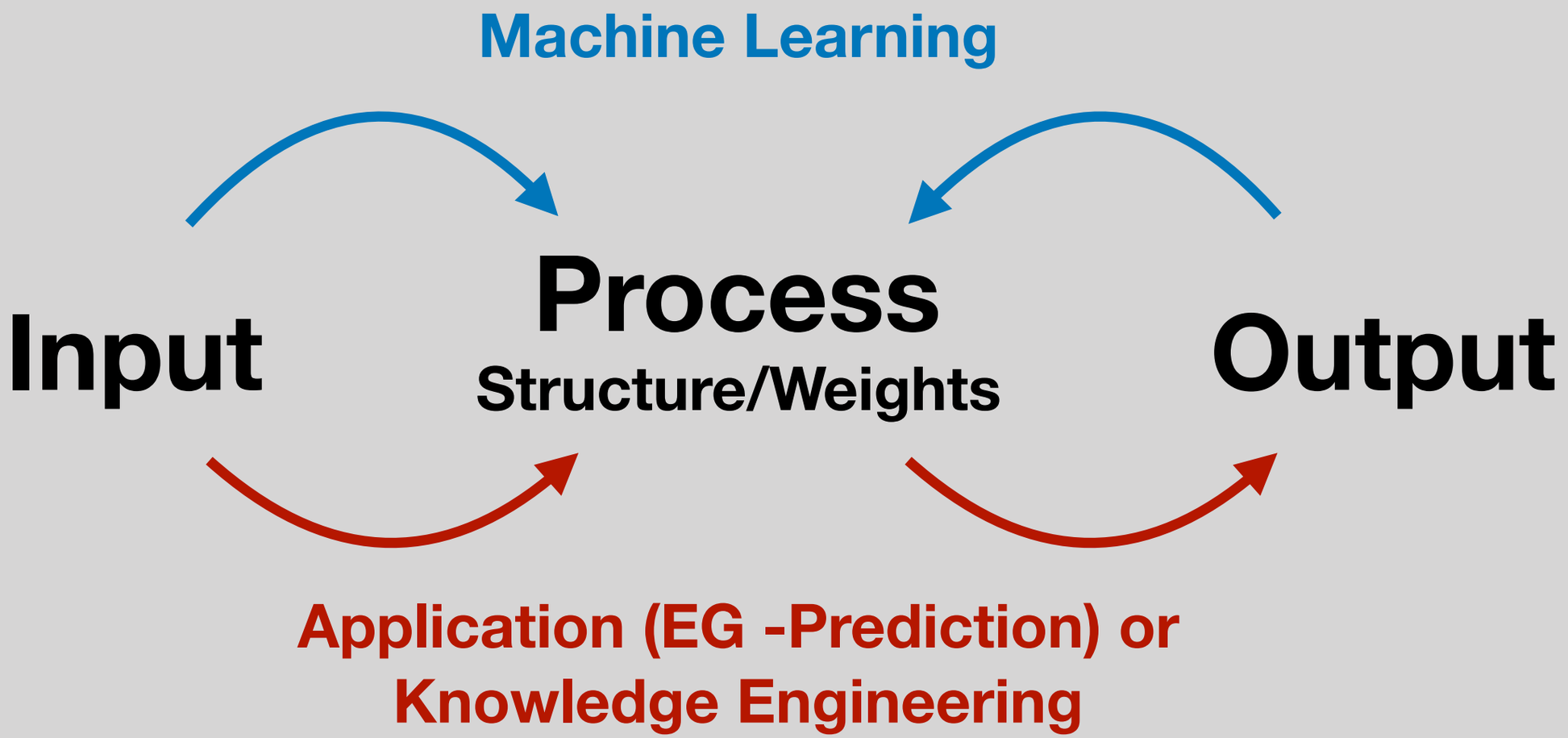
HUDK 5053: Feature Engineering Studio

10/11/18 11:12 AM

The Five Tribes of Machine Learning



ML Recap


Machine Learning
Input
Process
Structure/Weights
Output
Application (EG -Prediction) or
Knowledge Engineering

- Symbolists
- Connectionists
- Evolutionaries
- Bayesians
- Analogizers

Symbolists

Symbolists

- Inverse deduction (deduction)

Socrates is human

……..?

Therefore Socrates is mortal

- Decision trees

Symbolists

- Manipulating symbols with an algorithm and choose the ones that work
- It is about learning “rules” that can be applied

Symbolists

Positives:

- Easy to interpret
- Fast
- Makes “sense”

Drawbacks:

- The number of possible inductions is practically infinite - need to be very specific about the problem space
- If the premise or conclusion are wrong it is over
- Overfitting
- Concepts are rarely cleanly defined: female/male, spam/non-spam - can’t incorporate grey areas

Connectionists

Connectionists

- Hebb's Rule: neuron's that fire together, wire together
- One concept = many neurons
- Sigmoid curve
- Backpropagation

Connectionists

Positives

- Can learn very complex data sets

Negatives

- Hyperspace is ~infinite, you will likely find a local minima
- Weights are not interpretable
- Can’t do adaptive reasoning (rule chaining)

Evolutionaries

Evolutionaries

- John Holland (first PhD in CS)
- Objective, program, fitness function, sex
- Selective breeding + immortality
- EG - Spam filter that looks at every word in an email
- Mostly work at the sub-routine level

Evolutionaries

Positives

- Combines neural nets with rule based system
- Maybe it can create any kind of machine?

Negatives

- No empirical reason to have the sex step
  - And maybe a reason not to (mixability)
- Is it the evolutionary nature or just brute force that leads to success?
- Needs a lot of computing power

Bayesians

Bayesians

$$P(\theta \mid \mathbf{D}) = P(\theta)\frac{P(\mathbf{D} \mid \theta)}{P(\mathbf{D})}$$

- Probabilistic (just grey areas)
- Conditional probabilities shrink the problem space
- Often we know the probabilities of the effects given causes, what we want is the probabilities of the causes given the effects (EG - medical diagnoses)
- Conditional independence assumption

Bayesians

Positives

- Computationally simple
- Empirically accurate
- Can handle ambiguity

Negatives

- Conditional independence assumption
- Susceptible to exponential blowup/Bayesian networks become intractable as variables ↑
- There is no true hypothesis = have to calculate everything
- Can’t generate new hypotheses on the fly

Analogizers

Analogizers

- Representation = your data
- Find the thing closest to the thing you are looking for: nearest neighbor
- EG - John Snow Cholera Map (1854)
- Collaborative Filters
- Support Vector Machines

Analogizers

Positives

- Fast and at one time accurate as Neural Nets for complex feature sets OTB
- Can do transfer learning
- High dimensional space works well

Negatives

- Can’t handle class overlap well
- Run time is dependent on data size
- Probabilities are generated by cross validation