Week 1, Video 6

Advanced Classifiers

Classification

There is something you want to predict (“the label”)

The thing you want to predict is categorical

- The answer is one of a set of categories, not a number

Previous methods discussed

kNN, kStar, Decision Trees, Decision Rules, Step Regression, Logistic Regression

Conservative algorithms that do not attempt to capture all of the variance in the data

Previous methods discussed

kNN, kStar, Decision Trees, Decision Rules, Step Regression, Logistic Regression

These algorithms usually do better in most educational data mining than less conservative algorithms

- In brief, educational data has lots of systematic noise

That Said

There are clear places for some of the less conservative algorithms

Support Vector Machines

Conducts dimensionality reduction on data space and then fits hyperplane which splits classes

- For example, a 346 dimensional cloud of data becomes a 7 dimensional cloud of data
  And then the data is split into 2 groups by a hyperplane cutting across the 7 dimensions

Creates very sophisticated models

Support Vector Machines

- Great for text mining
- Great for sensor data
- Not commonly used with most other types of educational data
  - Logs, grades, interactions with software

Random Forest

Split the training data into random subsets of data points
For each subset, take a random set of the data features
Build a decision tree on the resultant data set

Take all the trees, each tree gets one vote

Neural Networks

Composes extremely complex relationships through combining “perceptrons”

Finds *very* complicated models

Soller & Stevens (2007)

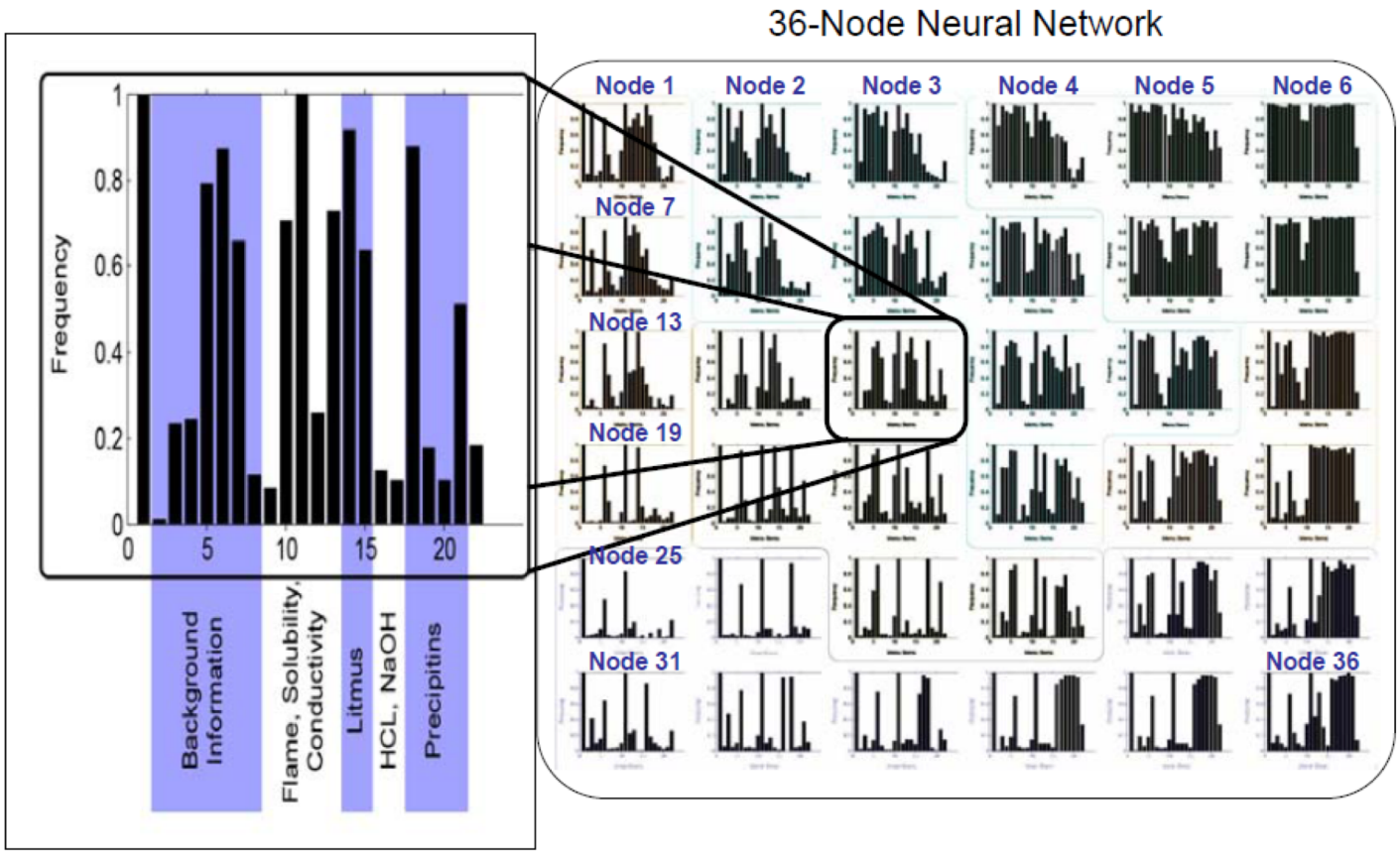

Figure 11. A Neural Network Showing the 36 Nodes, Each Describing a Different Subset of the Population

Recurrent Neural Networks

Also referred to as “deep learning”

Try to fit sequences of data rather than single data points

Neural Network with partial propagation of information over time

Several variants depending on how information propagates, such as long short-term memory networks (LSTMs)

Recurrent Neural Networks

Commonly used to represent language over time, with sentences and paragraphs represented as sequences of words

Recurrent Neural Networks

Used in the somewhat controversial deep knowledge tracing, discussed in week 4

- Measure of change of knowledge over time

Used in new detectors of student emotion, thought in initial analyses to be significantly more accurate than previous models (Botelho et al., 2017)

- Leveraging reliable trends and patterns in emotion over time

Recurrent Neural Networks

Used in models predicting MOOC dropout, which has a strong temporal character (Whitehill et al., 2017)

Used to predict humans' selections of dialogue moves in tutorial dialogue (Min et al., 2016)

A clear future for RNNs and their variants

- Still not clear where they will be most useful in education

Week One Complete!

Week Two

How do we know if a prediction model is any good?

- Goodness Metrics
- Model Validation