Finding Your WAY, Courtesy of Machine Learning

Chaitanya Asawa[1], Marcus Gomez[1], Viraj Mehta[2]

[1] Department of Computer Science, Stanford University, [2] Department of Mathematics, Stanford University

Motivation

Say you would like to fulfill your Formal Reasoning WAY with "Complex Analysis, Geometry, and Topology." Unfortunately, you can't.

Background

- Undergraduates at Stanford must take 11 courses across 8 different WAYS
- Currently, many Stanford courses are not assigned a WAY yet, even when they seemingly should

Problem Statement

- Given a course description, we would like to predict what WAY(S) it satisfies
- This is a multi-label, multi-class problem.
- Each WAY represents an output class
- Each course can satisfy a single or multiple WAY(S), so there are multiple labels

Dataset

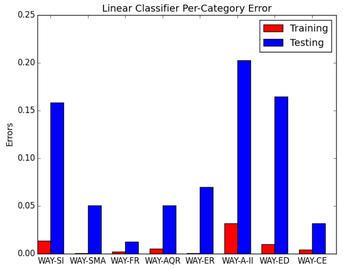
- We received access to ExploreCourses data
- We lowercased all words, removed stop words, and punctuation, and stemming in some cases
- Out of 14336 courses gathered, only **1571** satisfied WAY(S) – very limited data for multiple labels and classes

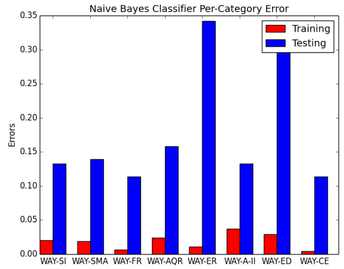
Evaluation Metric: Hamming Error

$$D_h^{(i)} = 1 - \frac{|A^{(i)} \cap A^{(i)}|}{|A^{(i)} \cup B^{(i)}|}$$

Baseline Models

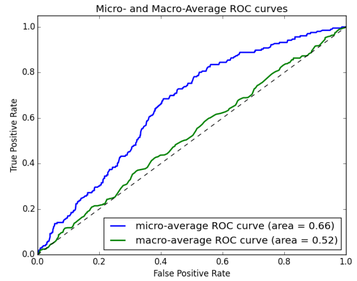
Linear and Naïve-Bayes OneVsRest classifiers using Word2Vec-style negative case subsampling



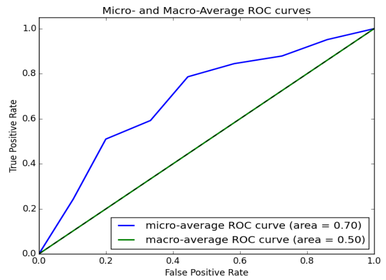


OneVsRest Support Vector Machines

Leveraged linear and Gaussian kernels



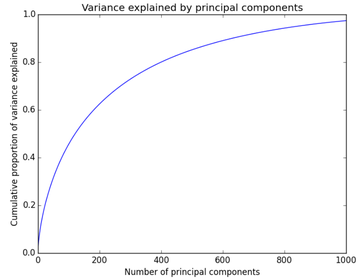
ROC curves for linear kernel SVM



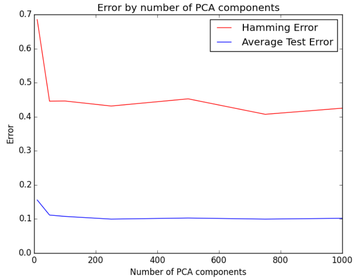
ROC curves for Gaussian kernel SVM

Dimensionality Reduction with PCA

Reduced model variance using first n principal components



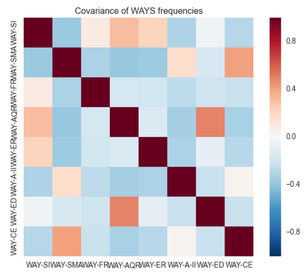
Percentage of variance explained by components by number of components used



Error of linear PCA as a function of number of components used

Independence

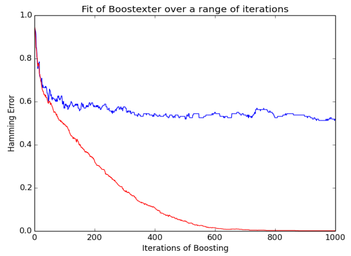
Linearity assumption is likely incorrect



Covariance of WAYS frequencies

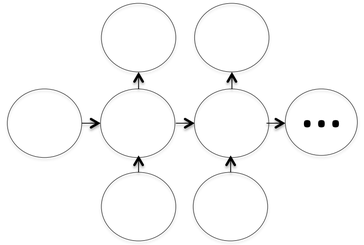
BoosTexter

Mod of AdaBoost for multi-class, multi-label problems



BoosTexter exhibits a classic overfitting curve

Neural Sequential Models

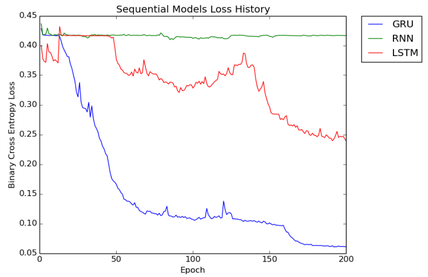


Generalized Recurrent Model

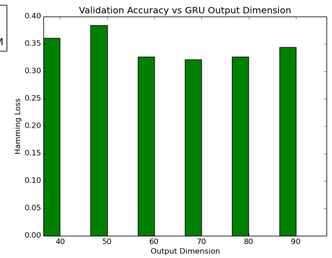
Hoping to capture nonlinear interactions between the words that our previous models may not have, we used deep learning sequential models, including a:

- RNN – Recurrent Neural Net
- LSTM – Long Short Term Memory Networks
- GRU – Gated Recurrent Unit

The inputs were course descriptions converted into a distributed vector representation using GloVe word vectors.

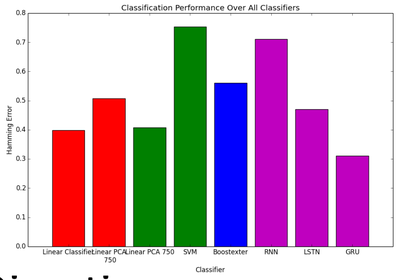


Performance of different neural models



Tuning GRU on output size does not have much effect

Comparison of All Models



Future Directions

We think an important issue was the dearth of data available to us, and believe that our models can perform better with more data.