Self-taught learning

Example: Classifying “Baseball” or “Football” webpages

- given only a few labeled examples of “Baseball” or “Football” webpages
- along with access to a large corpus of unrelated documents.

Raina et al (ICML’07) applied “sparse coding” to this problem.

- learn a sparse, higher-level representation from the unlabeled data, and apply the new representation to classification task.

\[
\text{input} = 0.8 \ast \text{features} + 0.3 \ast \text{features} + 0.5 \ast \text{features}
\]
Our approach: Exponential family sparse coding

- **Works well for self-taught learning with text classification.**
  - 9-23% average error reduction over raw bag-of-words features.
  - Better accuracy than (Gaussian) sparse coding, Exponential family PCA, and Latent Dirichlet Allocation (LDA).

- **An efficient algorithm** for general L1-regularized convex optimization problems.
  - Iteratively solves quadratic approximations with L1-penalty.
  - 2-20x faster for moderate sized problems with sparse solutions compared to state-of-the-art algorithms. (Lee et. al (AAAI’06), Koh et. al (JMLR’07), Andrew & Gao (ICML’07), Yu et. al (ICML’08)).