

CS276B Final Exam Practice Questions

1. Clustering. Consider the 3-means algorithm on a set S consisting of the following 6 points in the plane: $a=(0,0)$, $b=(8,0)$, $c=(16,0)$, $d=(0,6)$, $e=(8,6)$, $f=(16,6)$. The algorithm uses the Euclidean distance metric to assign each point to its nearest centroid; ties are broken in favor of the centroid to the left/down. A starting configuration is a subset of 3 starting points from S that form the initial centroids. A 3-partition is a partition of S into 3 subsets; thus $\{a,b,e\}$, $\{c,d\}$, $\{f\}$ is a 3-partition; clearly any 3-partition induces a set of three centroids in the natural manner. A 3-partition is stable if repetition of the 3-means iteration with the induced centroids leaves it unchanged.
 - a. How many starting configurations are there?
 - b. What are the stable 3-partitions?
 - c. What is the number of starting configurations leading to each of the stable 3-partitions in (b) above?
 - d. What is the maximum number of iterations from any starting configuration to its stable 3-partition?
2. Text Classification question. Consider a generative model in continuous 1-dimensional feature space in which data are generated from a class according to a Normal(μ , σ) distribution. Assume that there are two classes, and that you classify new data to the most likely class. Assume that you are given the prior probability of a class $P(C)$.
 - a. What is the discriminant function that is implied by this generative model and classification rule?
 - b. Give a discriminative model and classification rule that could be used instead to classify new instances in this feature space.
3. List each of the classifiers we have discussed in class. When does each one work well, and when does it fail? What can you say in general about the suitability of classifiers for text classification?
4. Give an example of a classifier with high variance and low bias. What are the advantages/disadvantages of classifiers with these qualities?
5. What is a kernel? How does the use of different kernels change the power of a support vector machine?
6. Adapt the general EM algorithm to the problem of lexicon construction. Define a parametric generative model for lexicon entries, and describe how you would perform the E step and the M step. How does this algorithm differ from the Riloff and Jones multilevel bootstrapping algorithm?
7. HMMs. Consider the following HMM probabilities for a 3 state HMM with states for Company, Person, and Background:

Emissions

Company Person Bkgrnd

Charles	0.1	0.3	0.1
Schwab	0.1	0.1	0.05
sneezed	0.01	0.01	0.1
(other)	0.79	0.59	0.75

Transitions

From\To	Company	Person	Bkgrnd
Company	0.3	0.1	0.6
Person	0.1	0.4	0.5
Bkgrnd	0.1	0.1	0.8

Assume that the start state probability vector is the same as the probabilities for transitioning to that state from the background state:

Start Probs

Company	0.1
Person	0.1
Bkgrnd	0.8

Now suppose the document is "Charles Schwab sneezed".

- a. What is the Viterbi state sequence for this document? Show your calculations, including the probability of this viterbi (state sequence, document) pair [i.e., $P(X,O|\Theta)$].
- b. What does this example show about what is required for HMMs to be used successfully for IE?

8. What are the assumptions made by each of the modules of the Harabagiu and Moldovan question answering system? How could each one be improved?

9. How would you automatically find synonymous gene names in a database of medical literature?

10. How could information extraction techniques be applied to improve the accuracy of first story detection?