

Question 1

When SEMISUPERVISED is false and FIXED_SEED is true, the second cluster seems to be the positive cluster with words like beautifully, touching, power, true and style. On subsequent runs, where we randomize the seed, the cluster's theme changes and the top words in each cluster are not very clear at showing the good/badness of each cluster. I think I am discovering different clusters on each iteration, but I think the clustering is identifying real patterns.

Question 2

1. bag of words: "introspective", "entertaining", "worth"
2. maybe bag of bigrams? "hard time"
3. mixed sentiments, hard to fix: "thrilling", "positively", "deceit", "tragedy"
4. bag of words: "aggressive", "manipulative"
5. mixed sentiments, hard to fix: "epic", "sincere", "crisis"
6. bag of words: "Trouble", "mess"
7. bag of words: "hate"
8. bag of words: "recommend"
9. need to parse and understand meaning, critic is saying his dreams should remain dreams, but dream has a positive sentiment
10. bag of words: "joy"

Question 3

There will be some problems extracting the sentiment by using a HMM. The word "a" or "an" works for "positively thrilling", "hard time", "manipulative whitewash", and "self-indulgent", but many of the reviews do not have this styling. Some of the reviews also start with sentiment phrases like "dizzily gorgeous", "impeccable", and "aggressive" which are pretty good indicators of the overall sentence sentiment.

Question 4 That sentence might be hard to parse for meaning because of the strange phrases that the critic uses to describe the film. Words like "meaty", "peppering" and "zingers" are not very common words and are unlikely to be picked up and interpreted correctly by a model. The usage of metaphors like "peppering the pages" makes it harder to parse for the full meaning even as a human.