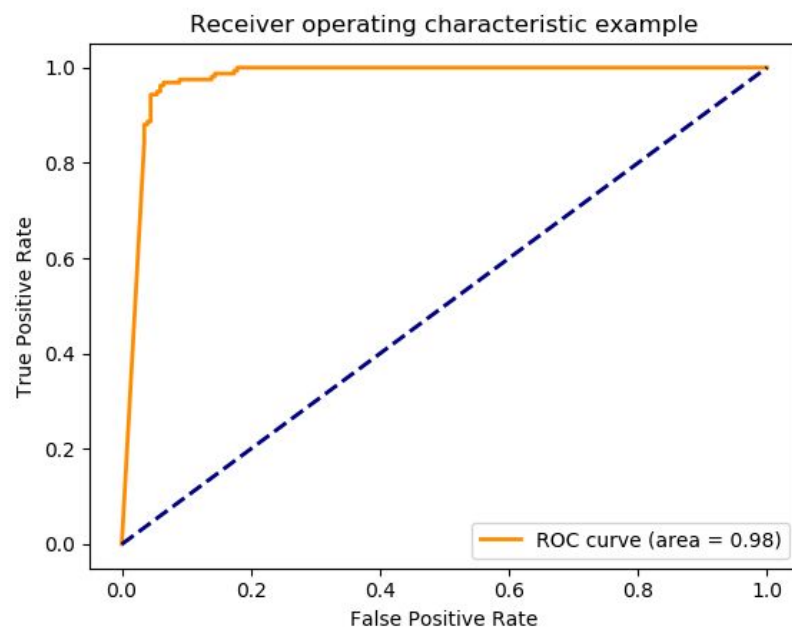


Naive Bayes

Study the [Naive Bayes](#) and [Sparse Matrix](#) classes. For the [messages dataset](#) build classifier. Plot [ROC curve](#) and output [confusion matrix](#). You can use [this example](#) to read messages.



1. Try modifying the message reader so that it also follows the message subject.
2. Try not default '*class_prior*' parameter such that no one legit message will be classified as spam but accuracy will be as high as possible.
3. Try using a 2-gram model (consider two consecutive words).