

1 Semi-Supervised Learning

1. (4pts) We want to leverage semi-supervised learning to classify online scientific documents to be "Fake" or "Real". Each document is represented using "bag of words" by assigning a binary indicator for the presence of 10,000 words in the document. For the following classifiers and learning methods discussed in class, state whether the method can be applied to improve the classifier (Yes) or not (No). Please provide a brief explanation. *Zero points for answers without correct justifications.*

(a) (2pts) Decision trees using re-weighting.

Solution: Yes. Re-weighting will impact the information gain for each feature.

(b) (2pts) Decision trees using EM.

Solution: No. Not a probabilistic method.