

2a.

The five words with the highest values of $f(j)$ are 'nations', 'free', 'war', 'government', and 'world'. The values of $f(j)$ for these words are 0.00318, 0.00331, 0.00372, 0.00394, and 0.00471 respectively. The five words with the lowest values of $f(j)$ are 'jobs', 'years', 'work', 'care', and 'americans'. Their values of $f(j)$ are -0.00526, -0.00383, -0.00358, -0.00342, and -0.00328 respectively.

2b.

I notice that the classifier is extremely confident regardless of whether it is right or wrong (with probabilities equal to 1 or very close to 1 in all cases). This may be caused by the fact that our model tries to get as many speeches as accurately labeled as possible, but it is not penalized for being confidently wrong and leads to the model overfitting.

2c.

One danger of a consistently confident classifier is that if it gets something incorrect, people will have to take action as though the model were correct (assuming the model is usually correct). This could lead to things such as patient injuries or proceeding with invasive procedures even when it may not be the best course of action. A classifier with confidence representing the probability it is correct would give more information to clinicians and other workers; this way they would know when to act upon a diagnosis and when to be more skeptical; this could be especially important if treatment is particularly expensive or comes with other unwanted side effects.

2d.

To introduce regularization, I might add more smoothing to our model. This would make small differences in features less important to the classifier such that it would be generally less confident in its predictions, only predicting things with high probability when they very clearly belong to a certain category.