Question 1

a) The weight vector is [suffix=ire, suffix=ton]. The model gets every single example wrong. Updating on all of these gives [0,0] → [−1,0] → [−1,−1] → [−1, 0] → [−1,−1], for a final answer of [−1,−1]
b) No, more or less any additional feature works that separates Washington including word length, prefixes ≥ 3, suffixes ≥ 4, etc.
c) Refer slides : perceptron can not solve XOR problem as perceptron is a linear model and linear model can not have a hypothesis which can provide linear segregation between output points in class 0 and class 1.

Question 2
a)The main modifications are to change WordEmbedding to CharEmbedding and to change the sentence splitting. For the DAN model, no other modifications are required.
b) Can be either smaller for speed or smaller for statistical efficiency
c) A few answers are possible here. Theoretically, the model could use information about characters to
accomplish something. However, this model is also extremely impoverished so it will not do well to
begin with.
d. The number of parameters should be less. First, the embeddings should be lower dimensional, which
leads to fewer parameters in the feedforward part of the DAN. Second, there are many fewer character embeddings than word embeddings. So the answer is the same regardless of whether you accounted for these embedding parameters or not.