Name:
NetID:

INFO 4940: Worksheet 00 (2/04)

1. Describe what the Softmax function does.

2. log (xy) = _____ + _____

3. Say you have a GPT-2-like language model with a 10,000 token vocabulary. If you pass this model a batch of 5 input sequences that are each 10 tokens long, what will be the shape of the model's logits?