CS 559B HW4 **Neil Gupte**

**Problem 1**

**As the networks use linear activation functions they essentially represent a linear regression model, in these models there is no use of stacking multiple hidden layers as no matter how many layers in the neural network, the last layer will be a linear function of the first layer. We cannot do backpropogation as the derivate is constant thus they are not useful for real world applications.**

**1)**

**The network A requires fewer resources and thus is computationally better. Cyclomaticaly better results.**

**It is also easier to implement.**

**2)**

**Network B can be used to learn compact representations, although network A is much better when using linear functions.**

**The network B can help in better generalization and is less prone to over fitting.**

**Problem 2**

1. **The classifier misclassifies only one point which is actually positive. So**

**ε= 1/8**

**α=**

**=**

**=**

**=**

**=0.972955**

**2) False**

**Not if the data in the training set cannot be separated by a linear combination of the specific type of weak classifiers we are using. For example consider the EXOR example with decision stumps as weak classifiers. No matter how many iterations are performed zero training error will not be achieved.**