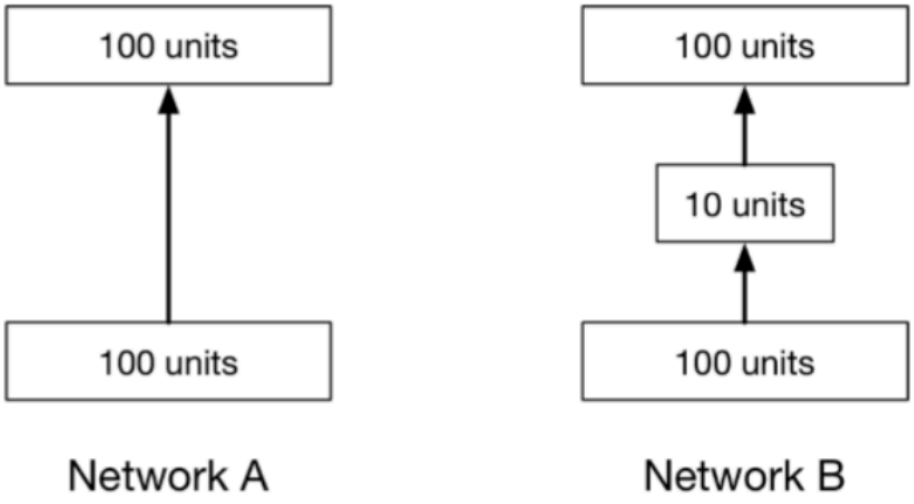
Problem 1 (50pt): Consider the following two neural networks, where all layers use *linear* activation functions. (Nonlinear activation, e.g., sigmoid, tanh etc, are NOT used in this problem) Hint: you may compare the number of parameters of these two networks and discuss.



(1) [25pt] Give one advantage of Network A over Network B.

(2) [25pt] Give one advantage of Network B over Network A.

1. Advantages of Network A over Network B:
    a. Easier to implement
    b. Fewer resources required (CPU, time etc.)
    c. A is more expressive than B

2. Advantages of Network B over Network A:
    a. B has a bottleneck layer, so the network is forced to learn a compact representation
    b. B has fewer connections, so backpropagation requires fewer operations
    c. B has fewer connections, so it's less prone to overfitting

Problem 2 (50pt): Adaboost

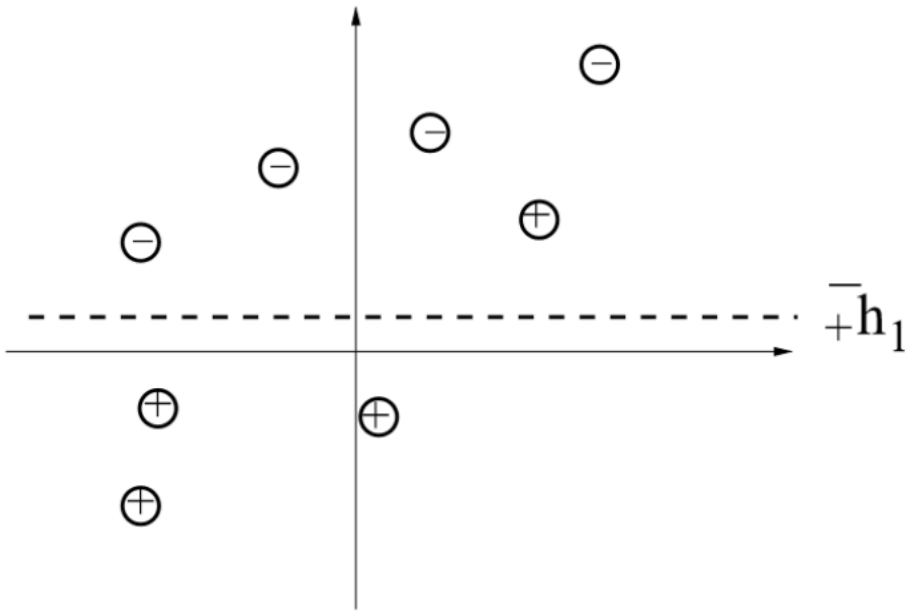


Figure 1: h_1 is chosen at the first iteration of boosting. Points above the h_1 are predicted to be negative while below the h_1 are predicted to be positive.

(1) [25pt] The above figure shows a dataset of 8 points, equally divided among the two classes (positive and negative). The figure also shows a particular choice of decision line h_1 picked by AdaBoost in the first iteration. What is the weight α_1 that will be assigned to h_1 by AdaBoost? (Initial weights of all the data points are equal, or $\frac{1}{8}$.)

(2) [25pt] AdaBoost will eventually reach zero training error, regardless of the type of weak classifier it uses, provided enough weak classifiers have been combined. (**True** or **False**, briefly explain)

1. $\alpha = \frac{1}{2}\log_2\frac{1-\epsilon}{\epsilon} = \frac{1}{2}\log_2\frac{7/8}{1/8} = \frac{1}{2}\log_2 7$
2. False

   If the data in the training set cannot be separated by a linear combination of the specific type of weak classifiers then it is not possible. For example, consider XOR like dataset.