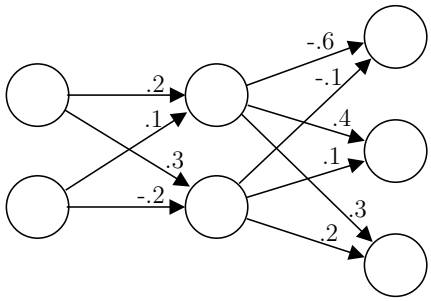
Neural Network Quiz

1. **Neural Network Mechanics**

   Consider the following three-layer neural network where each edge is labeled with the associated weight. (You may assume that all bias weights are zero.) The activation function at the hidden units is ReLU and the activation function at the output units is SoftMax.



   This network is intended to solve a three-class classification task for diabetes diagnosis. The three classes are:

   - Normal (top output unit)
   - Pre-Diabetes (middle output unit)
   - Diabetes (bottom output unit)

   The two inputs represent normalized results of an A1C test and a fasting blood sugar test.

   (a) Calculate the output for this network given that the input is $\mathbf{x} = [.1, -.4]$.

   (b) What is the cross-entropy loss, given that the true class label for this input is "Normal".

   (c) Write the weights for the hidden layer and output layer in matrix form.

2. **Counting Weights**

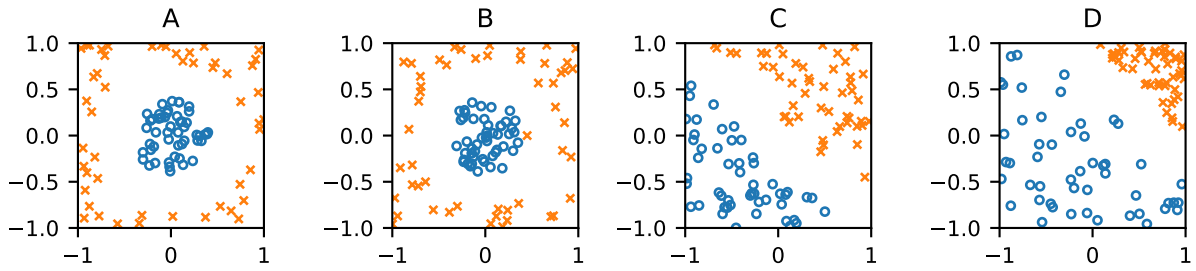
Consider a fully-connected neural network with the following configuration:

- Layer 1 (Input) : 100 units
- Layer 2 (Hidden): 10 units
- Layer 3 (Hidden): 10 units
- Layer 4 (Output): 4 units

How many overall weights does this network have, including bias weights?

3. **Classifier Capabilities**

Consider the following four datasets:



Fill in the following table indicating whether The indicated classifiers are capable of learning a model that will achieve 100% accuracy when trained on the datasets above. (Assume that the KNN classifiers are tested using leave-one-out cross validation.)

	A	B	C	D
Logistic Regression				
Broken Logistic Regression*				
Three-Layer Neural Network				
Decision Tree (max depth = ∞)				
Decision Tree (max depth = 4)				
Decision Tree (max depth = 1)				
1-Nearest Neighbor				
3-Nearest Neighbors				

*A logistic regression classifier with a broken bias weight stuck at 0.

4. **Mini-Batches**

What is a mini-batch? What is the benefit of using mini-batches in neural network training?

5. **Regularization**

(a) The following formula describes a neural network loss function with L2 regularization: $\mathbf{w}$ represents the full set of weights, L represents the unregularized loss function and $\|\mathbf{w}\|_2^2$ represents the squared L2 norm (the subscript indicates that this is an L"2" norm, the superscript represents that fact that we are squaring the result.)

$$L_\lambda(\mathbf{w}) = L(\mathbf{w}) + \lambda\|\mathbf{w}\|_2^2$$

- What will be the result of training a network with a very high value for λ?

- What will be the result of training a network with $\lambda = 0$?

- How would we go about selecting an appropriate value for λ?

(b) Dropout is sometimes explained in the context of bagging. Briefly explain the relationship between the two.