1. Which of the following is NOT true about the MNIST dataset in the lecture? (time stamp: 1:24)
   - ◯ Handwritten digits were extracted from cheques.
   - ◯ Can be loaded in Flux via `Flux.Data.MNIST.labels()`/`images()`.
   - ◯ Consists of RGB images of dimensions (28,28,3)
   - ◯ Contains a train:test split of 5000:55000 images.

2. Which of the following is NOT a function of defining a test set of 5000 images? (time stamp: 5:10)
   - ◯ The relatively small testset can reduce the training time of each epoch.
   - ◯ The testset helps estimate the model's generalisation capabilities.
   - ◯ The testset tests for signs of model overfitting (test accuracy $\ll$ train accuracy)
   - ◯ The testset will be integrated into the trainset at later epochs for finetuning.

3. What is the output of the softmax function in `model = Chain(Dense(n_inputs, n_outputs, identity)`, `softmax)`? (time stamp: 7:37)
   - ◯ 1 discrete label from the 0-9.
   - ◯ A 10 by 10 feature map.
   - ◯ 10 probabilities representing the likelihood that the input corresponds to each class.
   - ◯ A `10,1` one-hot vector, where the 1-value corresponds to the predicted class.

4. Which of the following is NOT true about the cross entropy (CE) function in `L(x,y)=Flux.crossentropy(model(x)`, `y)`? (time stamp: 7:53)
   - ◯ It calculates calculate the loss of the model.
   - ◯ It compares probability distributions.
   - ◯ It is equivalent to the log loss function.
   - ◯ It is a convex function.

5. What do a,b,c refer to respectively in `Flux.train!(a, [b], c)`? (time stamp: 8:52)
   - ◯ CE loss — training batch — SGD optimiser
   - ◯ training set — CE loss — SGD optimiser
   - ◯ training set — CE loss — Adam optimiser
   - ◯ Nesterov optimiser — CE loss — training batch

6. How is the throttle function applied to Flux callbacks? (time stamp: 10:30)
   - ◯ To pause the model training.
   - ◯ To update the learning rate automatically.
   - ◯ To redirect Jupyter Notebook output from stdout to a log file.
   - ◯ To control how frequently the callback function is executed.

7. Which is NOT a possible cause of having a test loss lower than a train loss? (time stamp: 13:55)
   - ◯ An overfitted model.
   - ◯ Test data being accidentally included in train set.
   - ◯ An incorrectly written loss function.
   - ◯ An underfitted model.

8. What does the number 20 represent in `model = Chain(Dense(n_inputs, 20, relu), Dense(20, n_outputs, identity), softmax)`? (time stamp: 18:55)

   ○ The total number of possible output classes.

   ○ The number of neurons in the hiden layer.

   ○ The number of hidden layers.

   ○ The spatial dimensions of input MNIST images.

9. Which of the following is NOT true about the ReLU function? (time stamp: 19:23)

   ○ It is a step function.

   ○ It is non-linear.

   ○ It increases model complexity.

   ○ It is a saturating function.

10. Which of the following is NOT a way to counter overfitting problems? (time stamp: 21:33)

   ○ Use more training data.

   ○ Use more test data.

   ○ Reduce the number of model parameters.

   ○ Shuffling data before defining train and test sets.

Answers: C, D, C, C, A, D, A, B, D, B