

Advanced Learning Algorithms

Week 1 to Week 4

- 1 Which of these are terms used to refer to components of an artificial neural network? (hint: three of these are correct)

Answer: [neurons](#); [layers](#); [activation function](#).

- 2 True/False? Neural networks take inspiration from, but do not very accurately mimic, how neurons in a biological brain learn.

Answer: [True](#).

- 3 For a neural network, what is the expression for calculating the activation of the third neuron in layer 2? Note, this is different from the question that you saw in the lecture video.

Answer: $a_3^{[2]} = g(w_2^{[2]}) \cdot a^{[1]} + b_3^{[2]}$.

- 4 For the handwriting recognition task discussed in lecture, what is the output $a_1^{[3]}$?

Answer: [The estimated probability that the input image is of a number 1, a number that ranges from 0 to 1.](#)

- 5 For the the following code:

```
model = Sequential([
    Dense(units=25, activation="sigmoid"),
    Dense(units=15, activation="sigmoid"),
    Dense(units=10, activation="sigmoid"),
    Dense(units=1, activation="sigmoid")])
```

This code will define a neural network with how many layers?

Answer: [4](#).

- 6 How do you define the second layer of a neural network that has 4 neurons and a sigmoid activation?

Answer: [Dense\(units = 4, activation = 'sigmoid'\)](#).

- 7 If the input features are temperature (in Celsius) and duration (in minutes), how do you write the code for the first feature vector x shown above?

Answer: [x = np.array\(\[\[200.0, 17.0\]\]\)](#).

- 8 According to the lecture, how do you calculate the activation of the third neuron in the first layer using NumPy?

Answer: [z13 = np.dot\(w13, x\) + b13, a13 = sigmoid\(z13\)](#).

- 9 According to the lecture, when coding up the numpy array W , where would you place the w parameters for each neuron?

Answer: [In the columns of \$W\$.](#)

- 10 For the code above in the "dense" function that defines a single layer of neurons, how many times does the code go through the "for loop"? Note that W has 2 rows and 3 columns.
Answer: 3 times.
- 11 For which type of task would you use the binary cross entropy loss function?
Answer: binary classification (classification with exactly 2 classes).
- 12 Which line of code updates the network parameters in order to reduce the cost?
Answer: `model.fit(X,y,epochs=100)`.
- 13 Which of the following activation functions is the most common choice for the hidden layers of a neural network?
Answer: ReLU.
- 14 For the task of predicting housing prices, which activation functions could you choose for the output layer? Choose the 2 options that apply.
Answer: ReLU; linear.
- 15 True/False? A neural network with many layers but no activation function (in the hidden layers) is not effective; that's why we should instead use the linear activation function in every hidden layer.
Answer: False.
- 16 For a multiclass classification task that has 4 possible outputs, the sum of all the activations adds up to 1. For a multiclass classification task that has 3 possible outputs, the sum of all the activations should add up to
Answer: 1.
- 17 For multiclass classification, the cross entropy loss is used for training the model. If there are 4 possible classes for the output, and for a particular training example, the true class of the example is class 3 ($y=3$), then what does the cross entropy loss simplify to? [Hint: This loss should get smaller when a_3 gets larger.]
Answer: $-\log(a_3)$.
- 18 For multiclass classification, the recommended way to implement softmax regression is to set from logits=True in the loss function, and also to define the model's output layer with
Answer: a linear activation.
- 19 The Adam optimizer is the recommended optimizer for finding the optimal parameters of the model. How do you use the Adam optimizer in TensorFlow?
Answer: When calling `model.compile`, set
`optimizer=tf.keras.optimizers.Adam(learning_rate=1e-3)`.
- 20 The lecture covered a different layer type where each single neuron of the layer does not look at all the values of the input vector that is fed into that layer. What is this name of the layer type discussed in lecture?
Answer: convolutional layer.

- 21 If the model's cross validation error J_{cv} is much higher than the training error J_{train} , this is an indication that the model has
 Answer: [high variance](#).
- 22 Which of these is the best way to determine whether your model has high bias (has underfit the training data)?
 Answer: [Compare the training error to the baseline level of performance](#).
- 23 You find that your algorithm has high bias. Which of these seem like good options for improving the algorithm's performance? Hint: two of these are correct.
 Answer: [Collect additional features or add polynomial features](#); [Decrease the regularization parameter lambda](#).
- 24 You find that your algorithm has a training error of 2%, and a cross validation error of 20% (much higher than the training error). Based on the conclusion you would draw about whether the algorithm has a high bias or high variance problem, which of these seem like good options for improving the algorithm's performance? Hint: two of these are correct.
 Answer: [Increase the regularization parameter lambda](#); [Collect more training data](#).
- 25 Which of these is a way to do error analysis?
 Answer: [Manually examine a sample of the training examples that the model misclassified in order to identify common traits and trends](#).
- 26 We sometimes take an existing training example and modify it (for example, by rotating an image slightly) to create a new example with the same label. What is this process called?
 Answer: [Data augmentation](#).
- 27 What are two possible ways to perform transfer learning? Hint: two of the four choices are correct.
 Answer: [You can choose to train just the output layers' parameters and leave the other parameters of the model fixed](#); [You can choose to train all parameters of the model, including the output layers, as well as the earlier layers](#).
- 28 Based on the decision tree shown in the lecture, if an animal has floppy ears, a round face shape and has whiskers, does the model predict that it's a cat or not a cat?
 Answer: [cat](#).
- 29 Take a decision tree learning to classify between spam and non-spam email. There are 20 training examples at the root node, comprising 10 spam and 10 non-spam emails. If the algorithm can choose from among four features, resulting in four corresponding splits, which would it choose (i.e., which has highest purity)?
 Answer: [Left split: 10 of 10 emails are spam. Right split: 0 of 10 emails are spam](#).
- 30 At a given node of a decision tree, 6 of 10 examples are cats and 4 of 10 are not cats. Which expression calculates the entropy $H(p_1)$ of this group of 10 animals?
 Answer: [-0.6log₂\(0.6\) - 0.4log₂\(0.4\)](#).

- 31 Before a split, the entropy of a group of 5 cats and 5 non-cats is $H(5/10)$. After splitting on a particular feature, a group of 7 animals (4 of which are cats) has an entropy of $H(4/7)$. The other group of 3 animals (1 is a cat) and has an entropy of $H(1/3)$. What is the expression for information gain?

Answer: $H(0.5) - (7/10H(4/7) + 3/10H(1/3))$.

- 32 To represent 3 possible values for the ear shape, you can define 3 features for ear shape: pointy ears, floppy ears, oval ears. For an animal whose ears are not pointy, not floppy, but are oval, how can you represent this information as a feature vector?

Answer: 0,0,1.

- 33 For a continuous valued feature (such as weight of the animal), there are 10 animals in the dataset. According to the lecture, what is the recommended way to find the best split for that feature?

Answer: Choose the 9 mid-points between the 10 examples as possible splits, and find the split that gives the highest information gain.

- 34 Which of these are commonly used criteria to decide to stop splitting? (Choose two.)

Answer: When the tree has reached a maximum depth; When the number of examples in a node is below a threshold.

- 35 For the random forest, how do you build each individual tree so that they are not all identical to each other?

Answer: Sample the training data with replacement.

- 36 You are choosing between a decision tree and a neural network for a classification task where the input x is a 100×100 resolution image. Which would you choose?

Answer: A neural network, because the input is unstructured data and neural networks typically work better with unstructured data.

- 37 What does sampling with replacement refer to?

Answer: Drawing a sequence of examples where, when picking the next example, first replacing all previously drawn examples into the set we are picking from.