机器学习课后题

Supervisied Learning

# 第一单元

Supervised vs unsupervised learning

有监督学习与无监督学习

1. Which are the two common types of supervised learning? (Choose two)
2. Clustering
3. Regression
4. Classification

Answer: B、C

1. Which of these is a type of unsupervised learning?
2. Regression
3. Classification
4. Clustering

Answer: C

# 第二单元

Regression

1. For linear regression, the model is

Which of the following are the inputs, or features, that are fed into the model and with which the model is expected to make a prediction?

1. and

Answer : A

1. For linear regression, if you find parameters and so that is very close to zero, what can you conclude?
2. This is never possible -- there must be a bug in the code.
3. The selected values of the parameters and cause the algorithm to fit the training set really poorly.
4. The selected values of the parameters and cause the algorithm to fit the training set really well.

Answer: B

# 第三单元

Train the model with gradient descent

使用梯度下降法训练模型

1. Gradient descent is an algorithm for finding values of parameters w and b that minimize the cost function J.

When is a negative number (less than zero), what happens to after one update step?

1. stays the same
2. decreases
3. increases.
4. It is not possible to tell if will increase or decrease.

Anawer : C

1. For linear regression, what is the update step for parameter b?

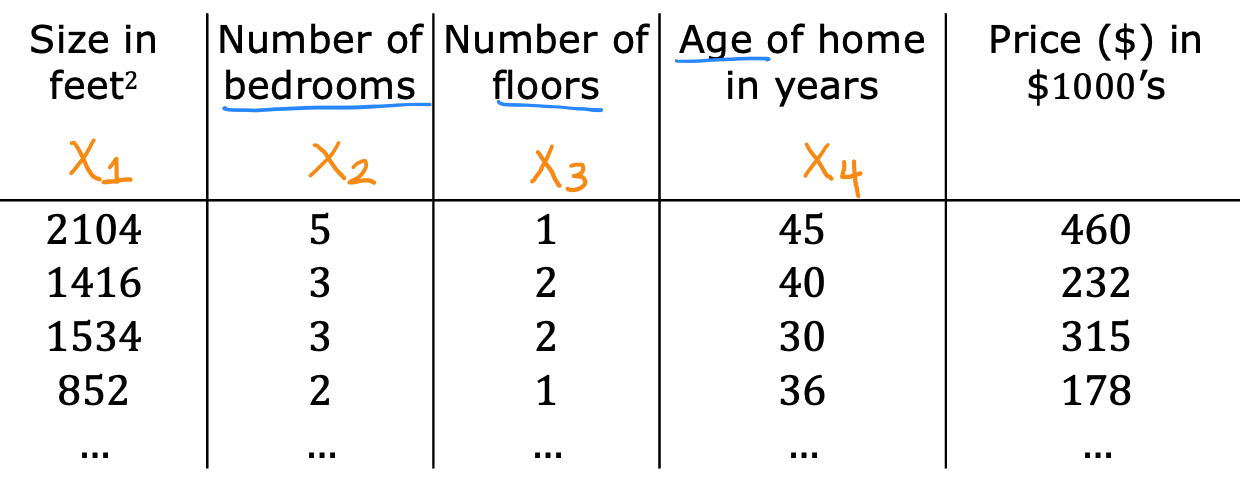
Answer: A

# 第一单元

Multiple linear regression

多变量线性回归

1. In the training set below, what is ？Please type in the number below (this is an integer such as 123, no decimal points).



Answer : 30

1. Which of the following are potential benefits of vectorization? Please choose the best option.
2. It makes your code run faster
3. It can make your code shorter
4. It allows your code to run more easily on parallel compute hardware
5. All of the above

Answer :D

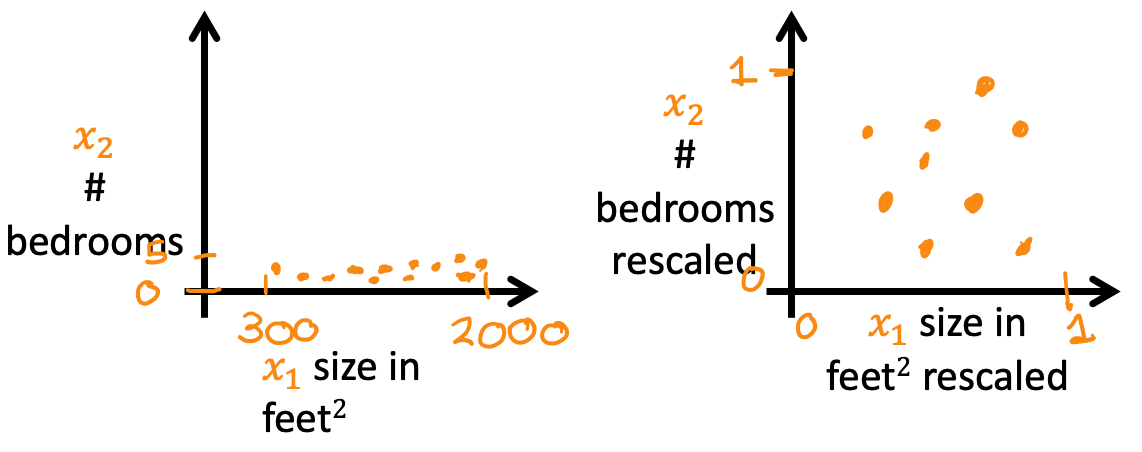
1. True/False? To make gradient descent converge about twice as fast, a technique that almost always works is to double the learning rate *alpha*.
2. False
3. True

Answer : A

# 第二单元

Gradient descent in practice

梯度下降的实践

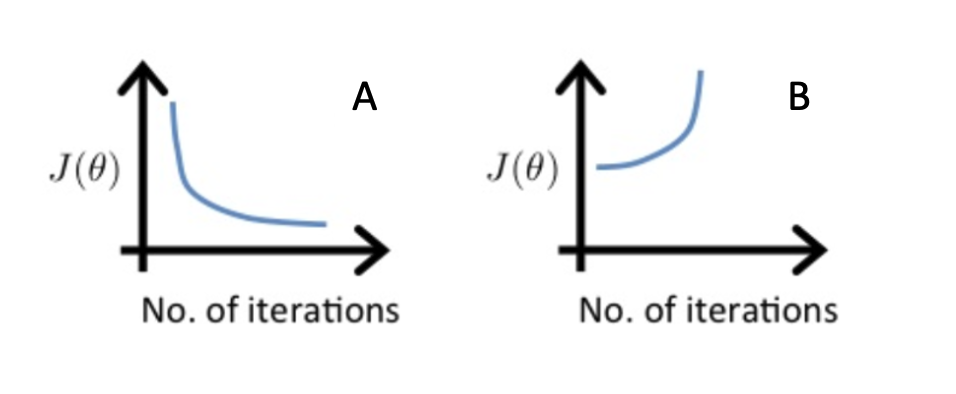


Which of the following is a valid step used during feature scaling?

1. Subtract the mean (average) from each value and then divide by the (max - min).
2. Add the mean (average) from each value and and then divide by the (max - min).

Answer : A

1. Suppose a friend ran gradient descent three separate times with three choices of the learning rate and plotted the learning curves for each (cost J for each iteration).



For which case, A or B, was the learning rate likely too large?

1. case B only
2. Neither Case A nor B
3. case A only
4. Both Cases A and B

Answer:A

1. Of the circumstances below, for which one is feature scaling particularly helpful?
2. Feature scaling is helpful when all the features in the original data (before scaling is applied) range from 0 to 1.
3. Feature scaling is helpful when one feature is much larger (or smaller) than another feature.

Answer:B

1. You are helping a grocery store predict its revenue, and have data on its items sold per week, and price per item. What could be a useful engineered feature?
2. For each product, calculate the number of items sold times price per item.
3. For each product, calculate the number of items sold divided by the price per item.

Answer:A

1. True/False? With polynomial regression, the predicted values does not necessarily have to be a straight line (or linear) function of the input feature x.
2. True
3. False

Answer:A

# 第一单元

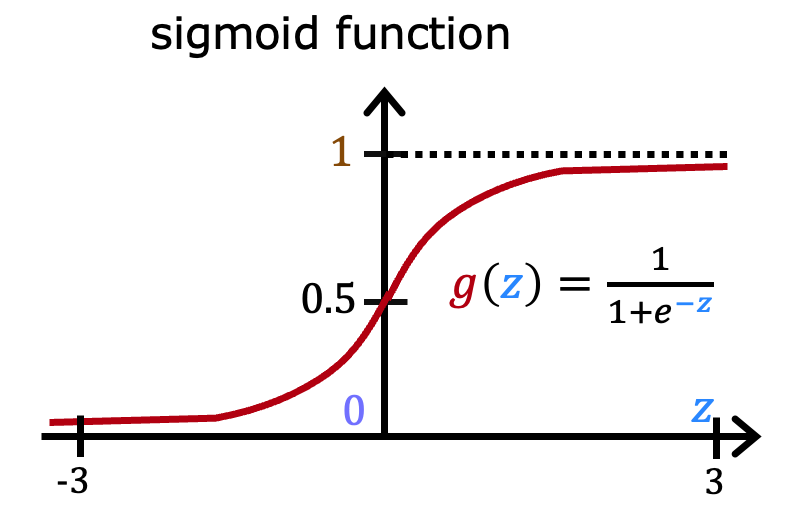
Classification with logistic regression

利用逻辑回归进行分类

1. Which is an example of a classification task?
2. Based on a patient's blood pressure, determine how much blood pressure medication (a dosage measured in milligrams) the patient should be prescribed.
3. Based on a patient's age and blood pressure, determine how much blood pressure medication (measured in milligrams) the patient should be prescribed.
4. Based on the size of each tumor, determine if each tumor is malignant (cancerous) or not.

Answer:B

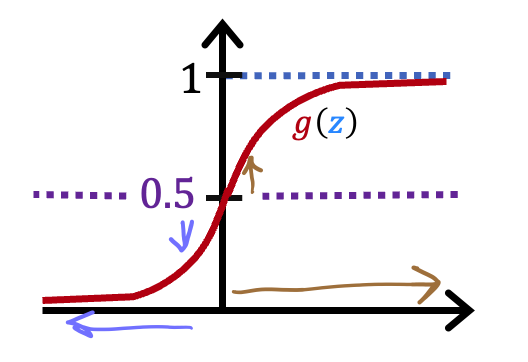
1. Recall the sigmoid function is



If z is a large positive number, then:

1. will be near 0.5
2. is near one (1)
3. will be near zero (0)
4. is near negative one (-1)

Answer:B



A cat photo classification model predicts 1 if it's a cat, and 0 if it's not a cat. For a particular photograph, the logistic regression model outputs (a number between 0 and 1). Which of these would be a reasonable criteria to decide whether to predict if it’s a cat?

1. Predict it is a cat if < 0.7
2. Predict it is a cat if < 0.5
3. Predict it is a cat if = 0.5
4. Predict it is a cat if >= 0.5

Answer:D

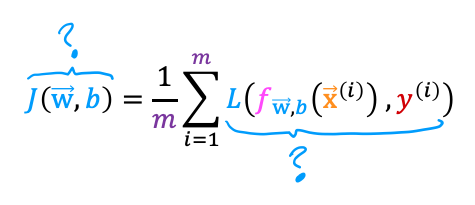
1. True/False? No matter what features you use (including if you use polynomial features), the decision boundary learned by logistic regression will be a linear decision boundary.
2. True
3. False

Answer:B

# 第二单元

Cost function for logistic regression

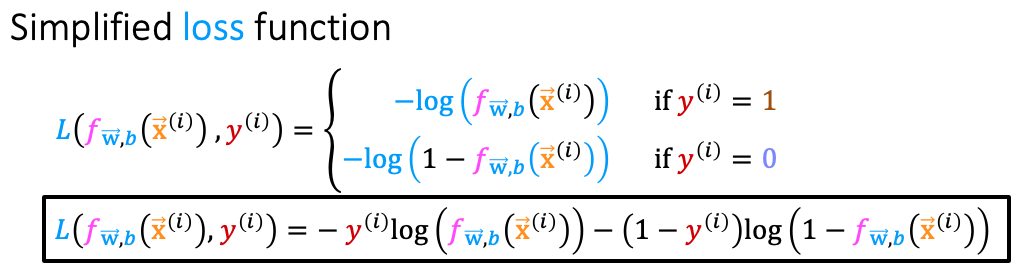
逻辑回归的代价函数



In this lecture series, "cost" and "loss" have distinct meanings. Which one applies to a single training example?

1. Loss
2. Cost
3. Both Loss and Cost
4. Neither Loss nor Cost

Answer:A



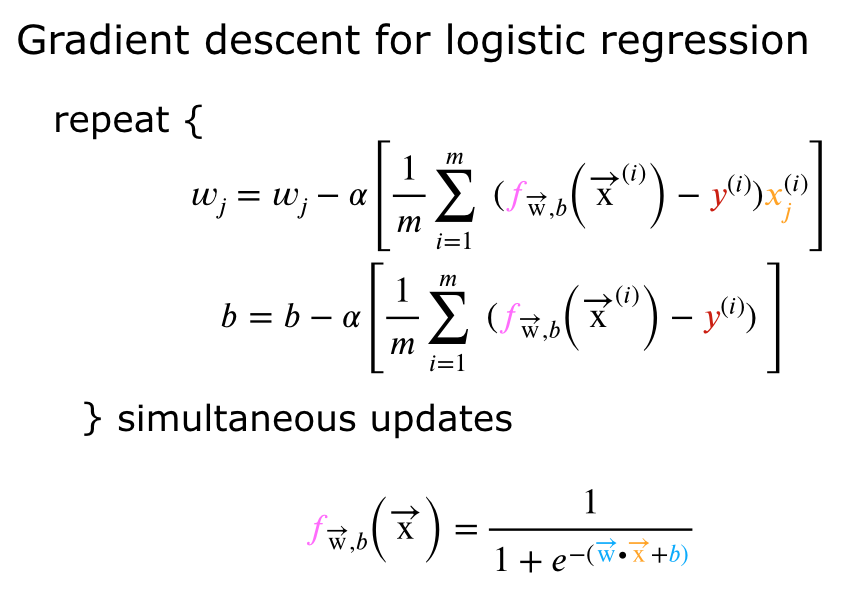
For the simplified loss function, if the label ，then what does this expression simplify to?

Answer:A

# 第三单元

Gradient descent for logistic regression

逻辑回归的梯度下降



Which of the following two statements is a more accurate statement about gradient descent for logistic regression?

1. The update steps are identical to the update steps for linear regression.
2. The update steps look like the update steps for linear regression, but the definition of is different.

Answer:A

# 第四单元

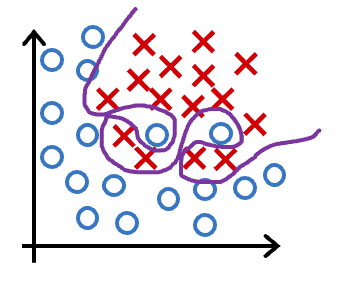
The problem of overfitting

过拟合问题

1. Which of the following can address overfitting?
2. Apply regularization
3. Select a subset of the more relevant features.
4. Remove a random set of training examples
5. Collect more training data

Answer: A、B、D

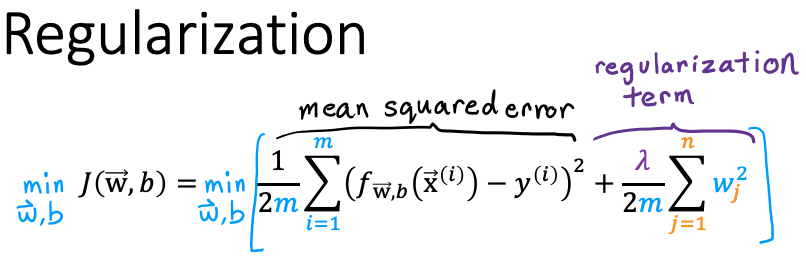
1. You fit logistic regression with polynomial features to a dataset, and your model looks like this.



What would you conclude? (Pick one)

1. The model has high bias (underfit). Thus, adding data is likely to help
2. The model has high bias (underfit). Thus, adding data is, by itself, unlikely to help much.
3. The model has high variance (overfit). Thus, adding data is, by itself, unlikely to help much.
4. The model has high variance (overfit). Thus, adding data is likely to help

Answer:C

  
Suppose you have a regularized linear regression model.  If you increase the regularization parameter , what do you expect to happen to the parameters ？

1. This will reduce the size of the parameters
2. This will increase the size of the parameters

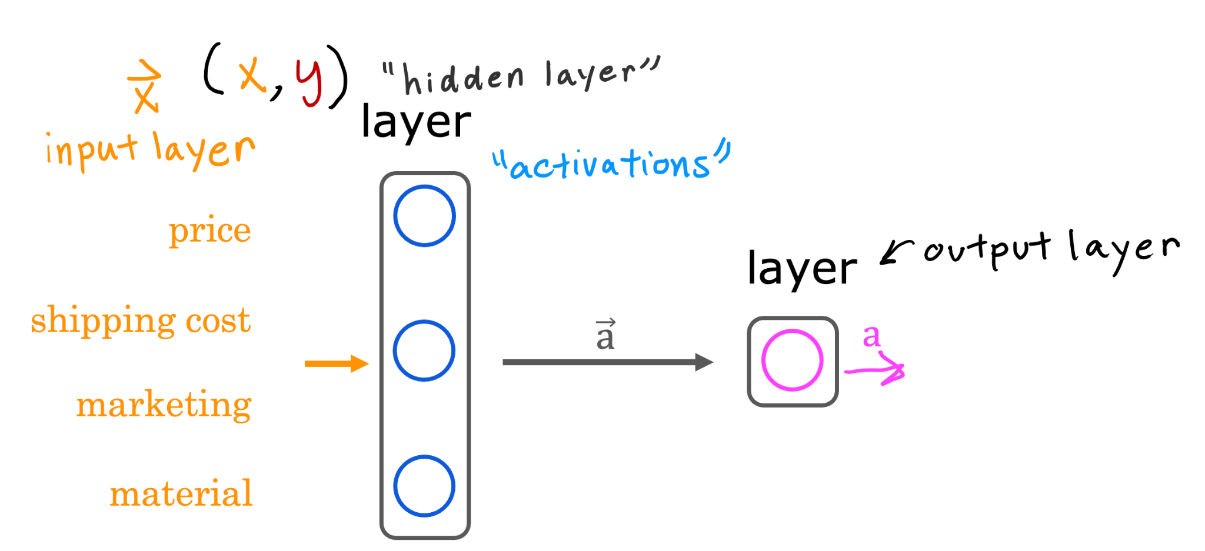
Answer:A

Advanced Learning

# 第一单元

Neural networks intuition

神经网络的直观了解



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

1. axon
2. neurons
3. activation function
4. layers

Answer: B、C、D

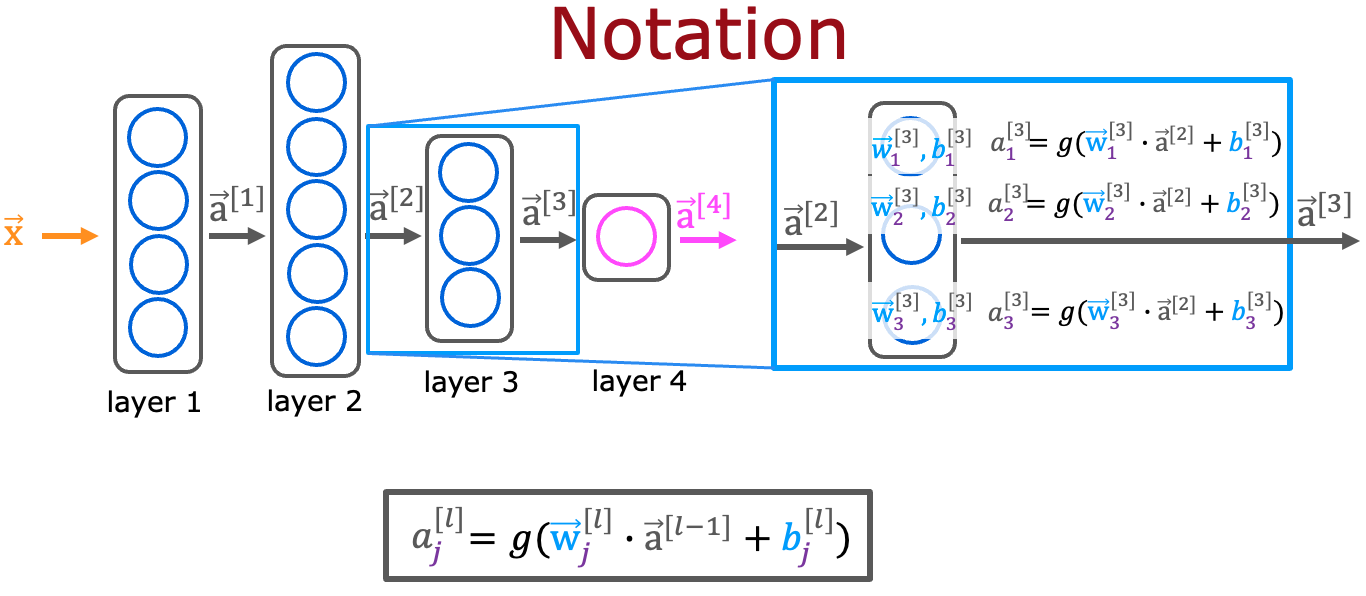
1. True/False? Neural networks take inspiration from, but do not very accurately mimic, how neurons in a biological brain learn.
2. True
3. False

Answer:A

# 第二单元

Neural network model

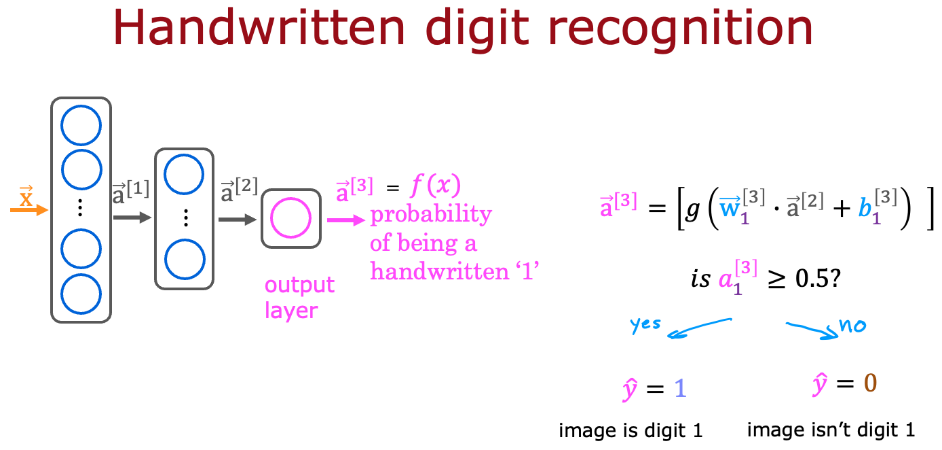
神经网络模型



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.

1. 

Answer:A



For the handwriting recognition task discussed in lecture, what is the output ？ B

1. A number that is either exactly 0 or 1, comprising the network’s prediction
2. The estimated probability that the input image is of a number 1, a number that ranges from 0 to 1.
3. A vector of several numbers that take values between 0 and 1
4. A vector of several numbers, each of which is either exactly 0 or 1

Answer:B

# 第三单元

TensorFlow implementation

TensorFlow 的实现

1. For the the following code:

model = Sequential([

Dense(units=25, activation="sigmoid"),

Dense(units=15, activation="sigmoid"),

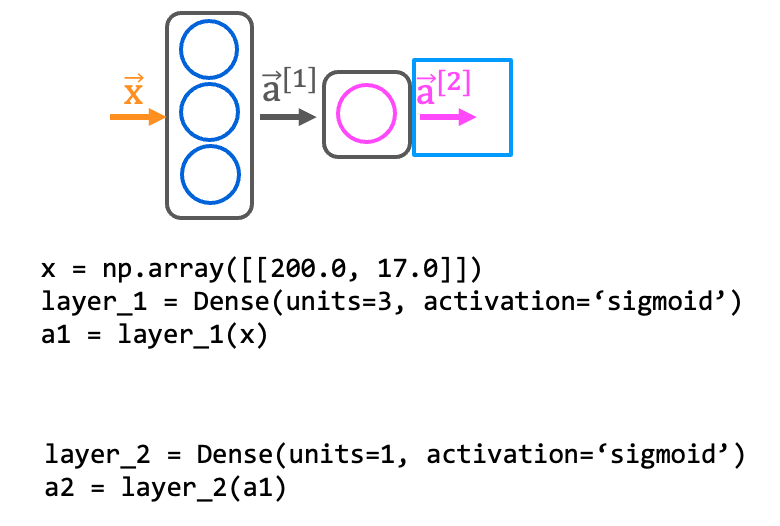
Dense(units=10, activation="sigmoid"),

Dense(units=1, activation="sigmoid")])

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

1. 5
2. 3
3. 4
4. 25

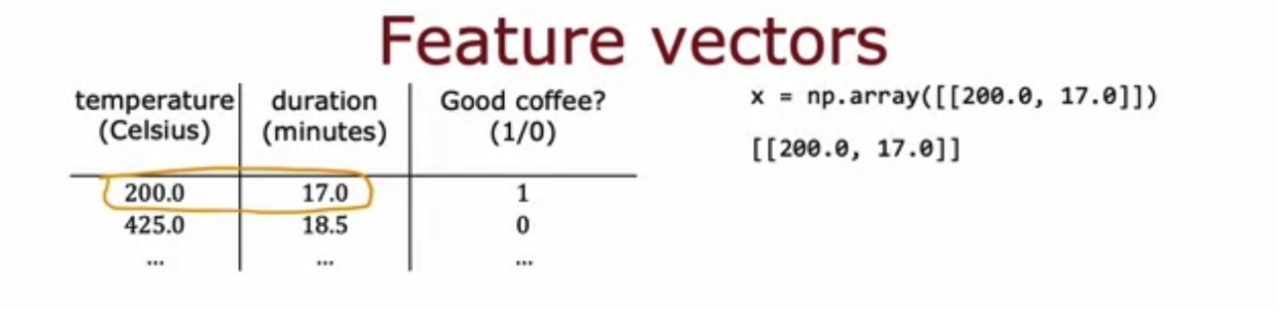
Answer:C



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

1. Dense(layer=2, units=4, activation = ‘sigmoid’)
2. Dense(units=4, activation=‘sigmoid’)
3. Dense(units=[4], activation=[‘sigmoid’])
4. Dense(units=4)

Answer:B



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?

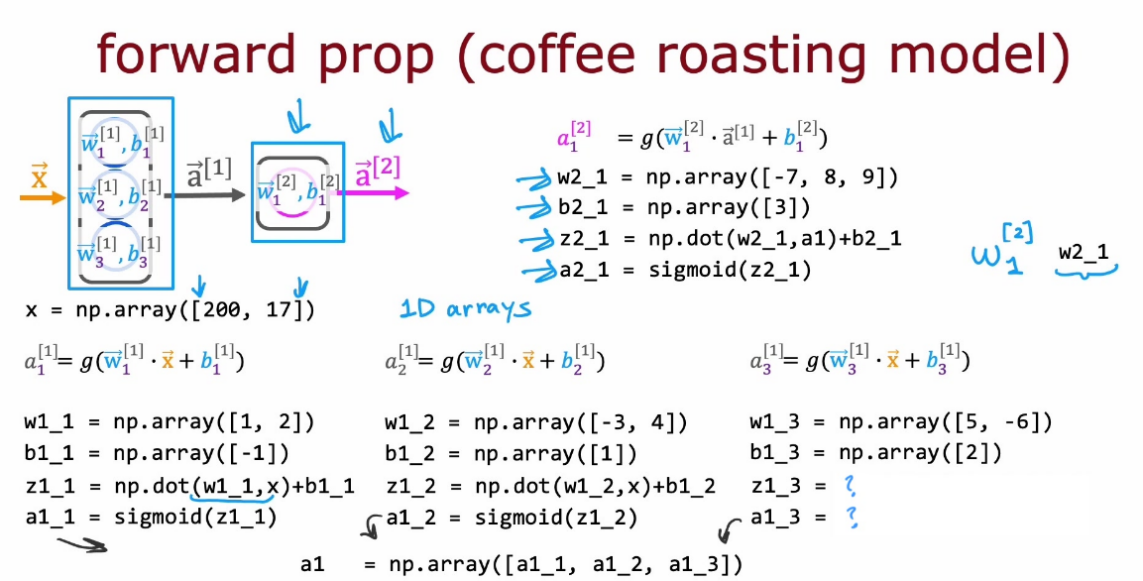
1. x = np.array([[200.0],[17.0]])
2. x = np.array([[200.0, 17.0]])
3. x = np.array([[‘200.0’, ’17.0’]])
4. x = np.array([[200.0 + 17.0]])

Answer:B

# 第四单元

Neural network implementation in Python

用 Python 实现神经网络



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

1. z1\_3 =w1\_3 \* x + b

a1\_3 = sigmoid(z1\_3)

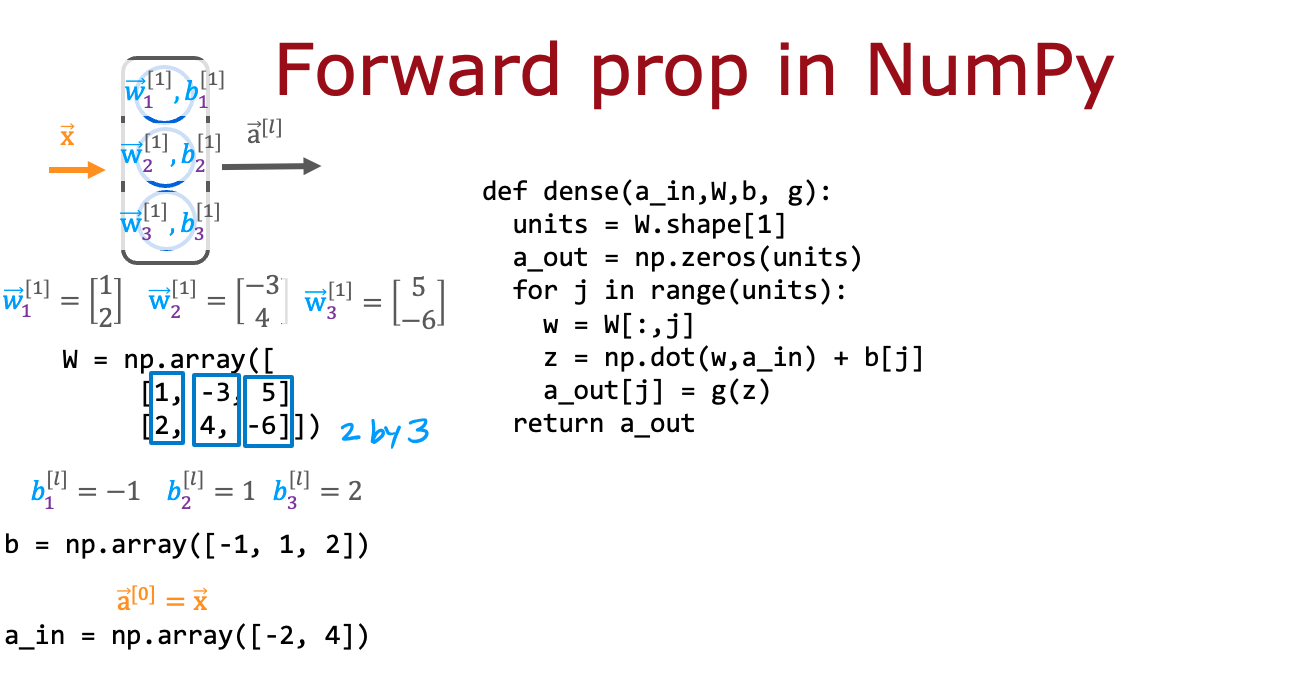
1. layer\_1 = Dense(units=3, activation='sigmoid')

a\_1 = layer\_1(x)

1. z1\_3 = np.dot(w1\_3, x) + b1\_3

a1\_3 = sigmoid(z1\_3)

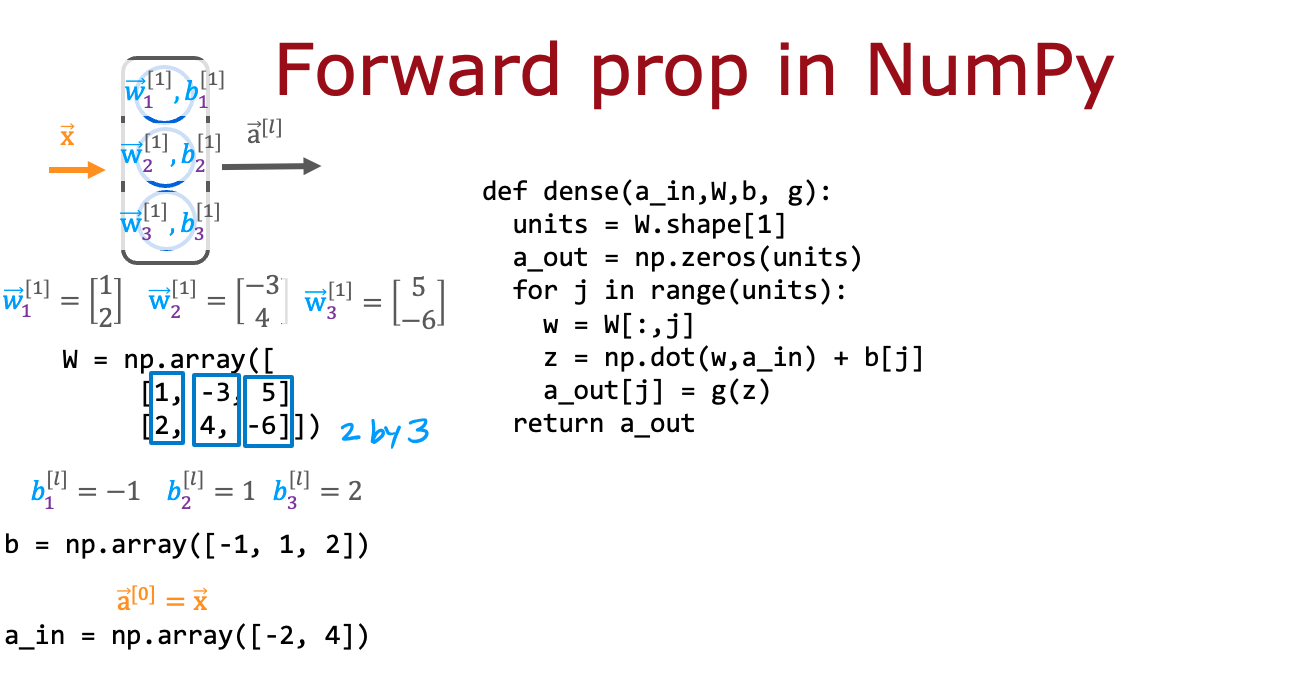
Answer:C



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

1. In the columns of W.
2. In the rows of W

Answer:A



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.

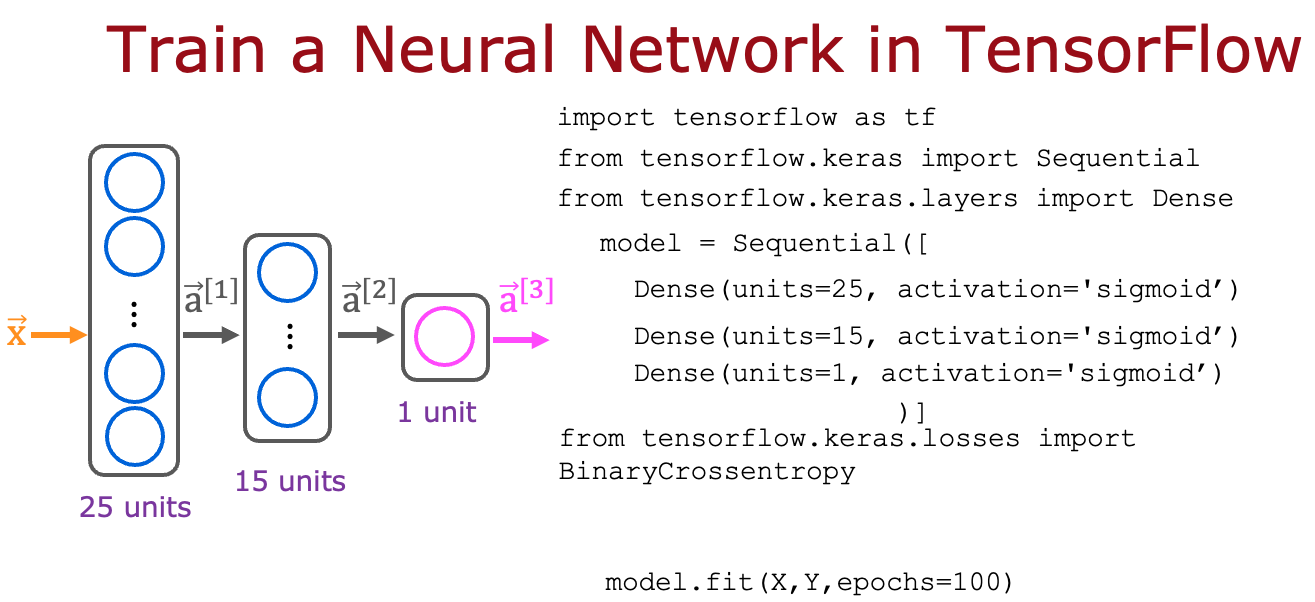
1. 6 times
2. 2 times
3. 3 times
4. 5 times

Answer:C

# 第一单元

Neural Network Training

神经网络的训练



Here is some code that you saw in the lecture:

```

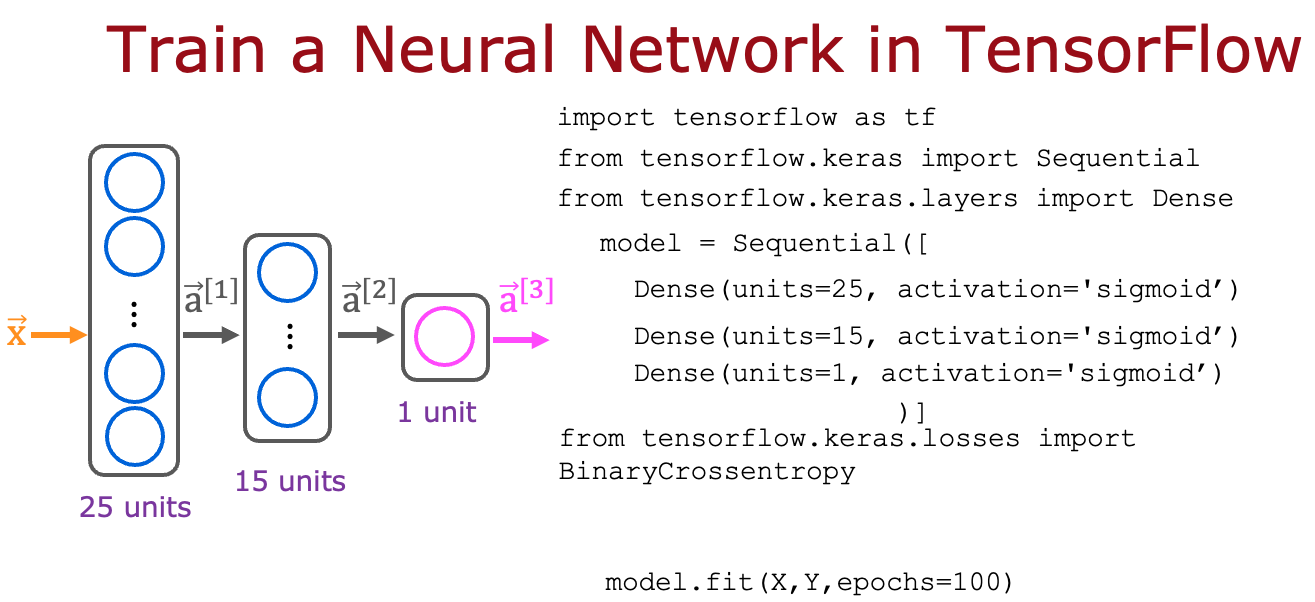
model.compile(loss=BinaryCrossentropy())

```

For which type of task would you use the binary cross entropy loss function?

1. BinaryCrossentropy() should not be used for any task.
2. A classification task that has 3 or more classes (categories)
3. regression tasks (tasks that predict a number)
4. binary classification (classification with exactly 2 classes)

Answer:B



Here is code that you saw in the lecture:

```

model = Sequential([

Dense(units=25, activation='sigmoid’),

Dense(units=15, activation='sigmoid’),

Dense(units=1, activation='sigmoid’)

])

model.compile(loss=BinaryCrossentropy())

model.fit(X,y,epochs=100)

```

Which line of code updates the network parameters in order to reduce the cost?

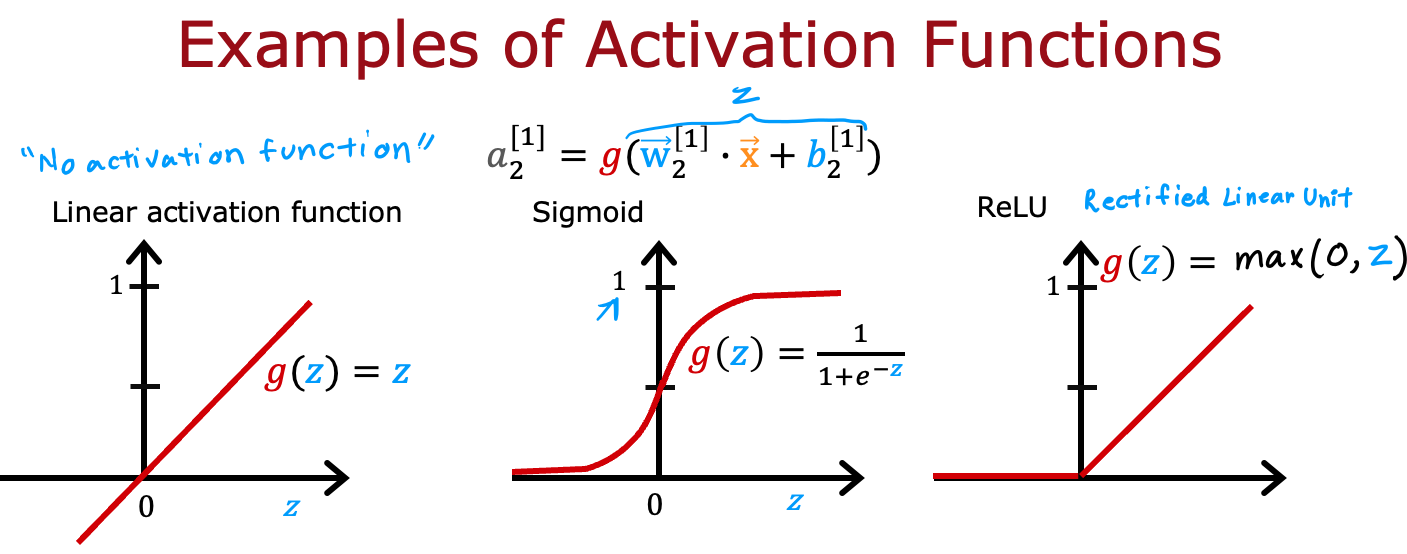
1. None of the above -- this code does not update the network parameters.
2. model.compile(loss=BinaryCrossentropy())
3. model = Sequential([...])
4. model.fit(X,y,epochs=100)

Answer:D

# 第二单元D

Activation Functions

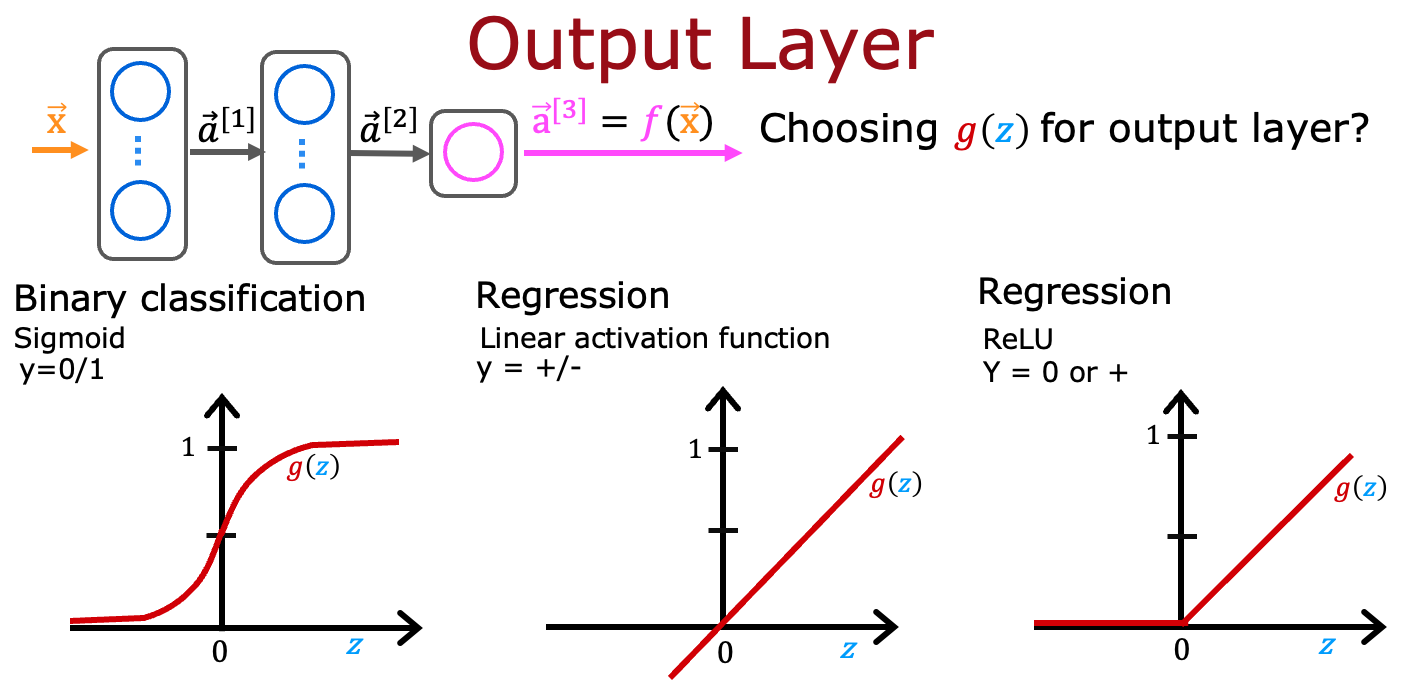
激活函数



Which of the following activation functions is the most common choice for the hidden layers of a neural network?

1. ReLU (rectified linear unit)
2. Linear
3. Sigmoid
4. Most hidden layers do not use any activation function

Answer:A



For the task of predicting housing prices, which activation functions could you choose for the output layer? Choose the 2 options that apply.

1. Sigmoid
2. linear
3. ReLU

Answer:B、C

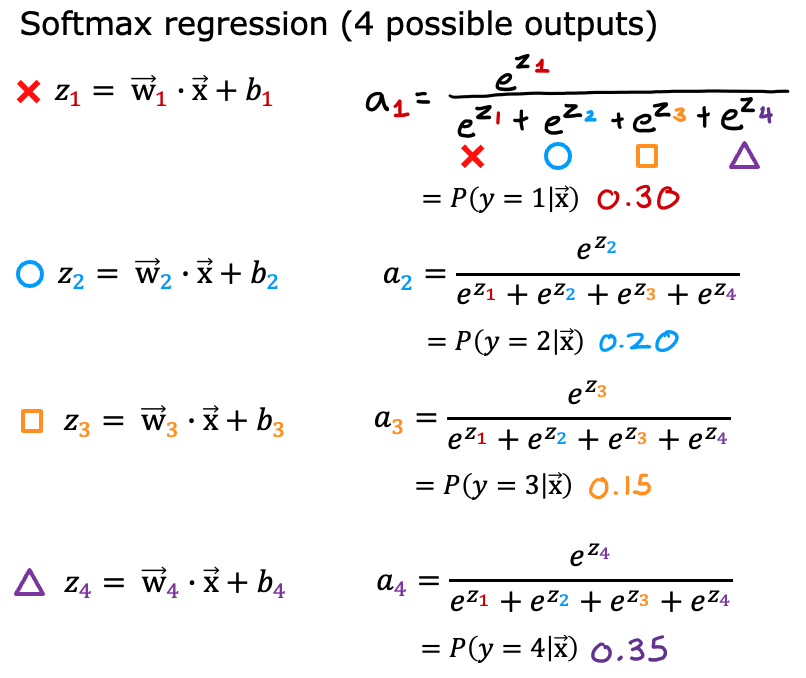
1. True/False? A neural network with many layers but no activation function (in the hidden layers) is not effective; that’s why we should instead use the linear activation function in every hidden layer.
2. True
3. False

Answer:B

# 第三单元

Multiclass Classification

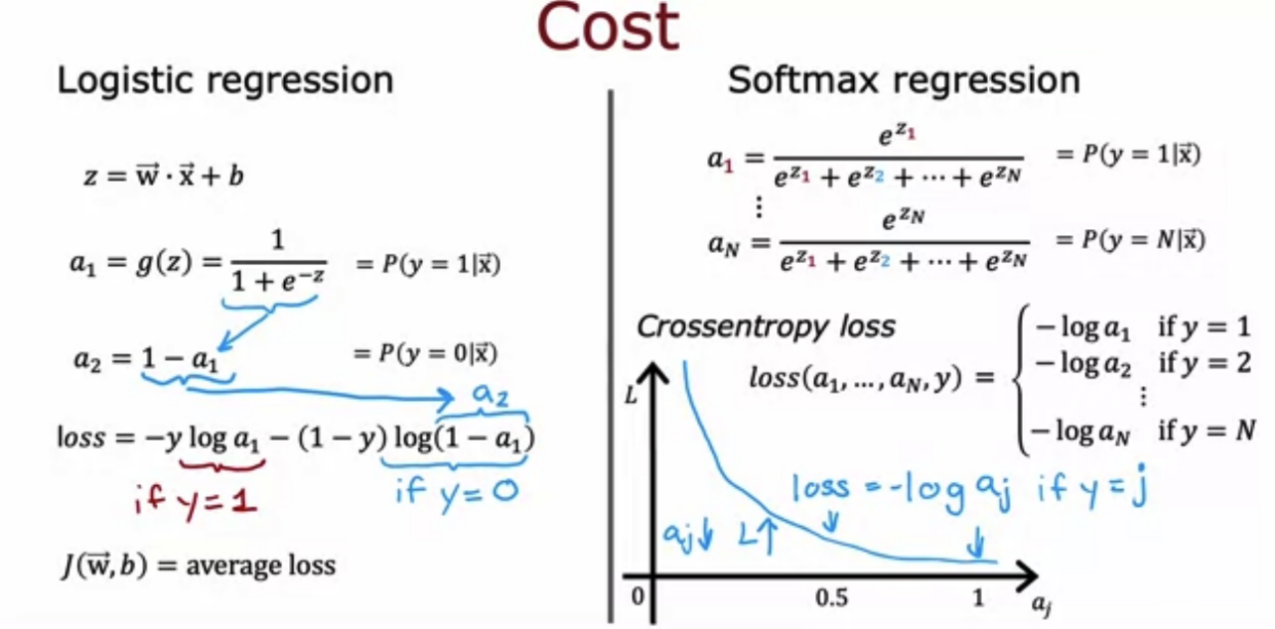
多类别分类



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 ….

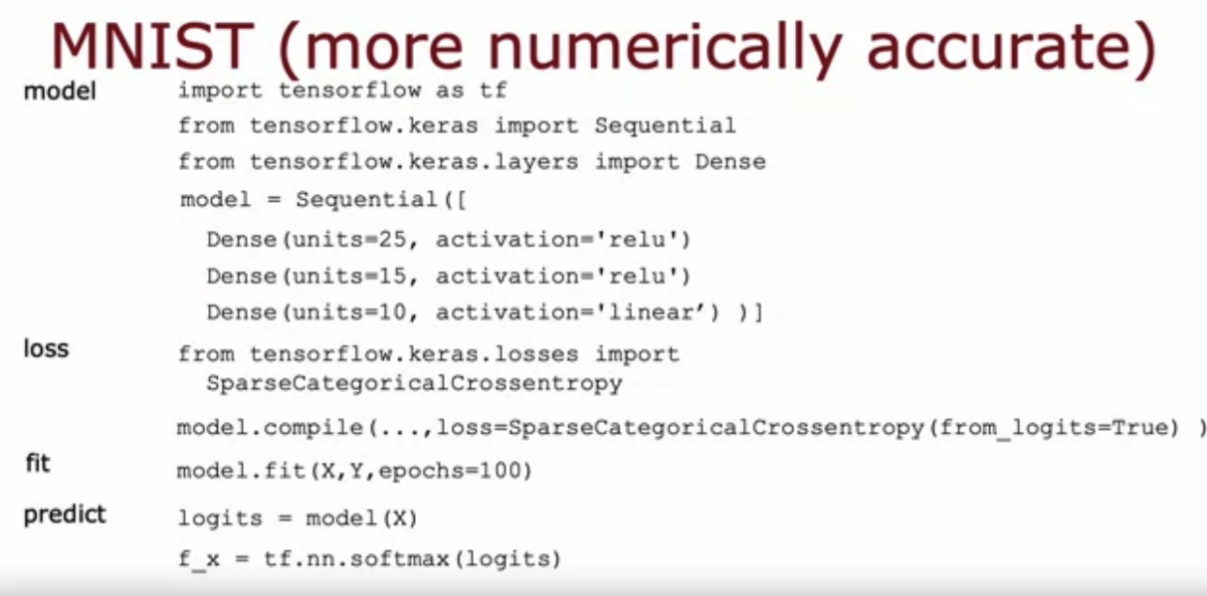
1. More than 1
2. It will vary, depending on the input x.
3. Less than 1
4. 1

Answer:D



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 gets larger.]

Answer:D



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…

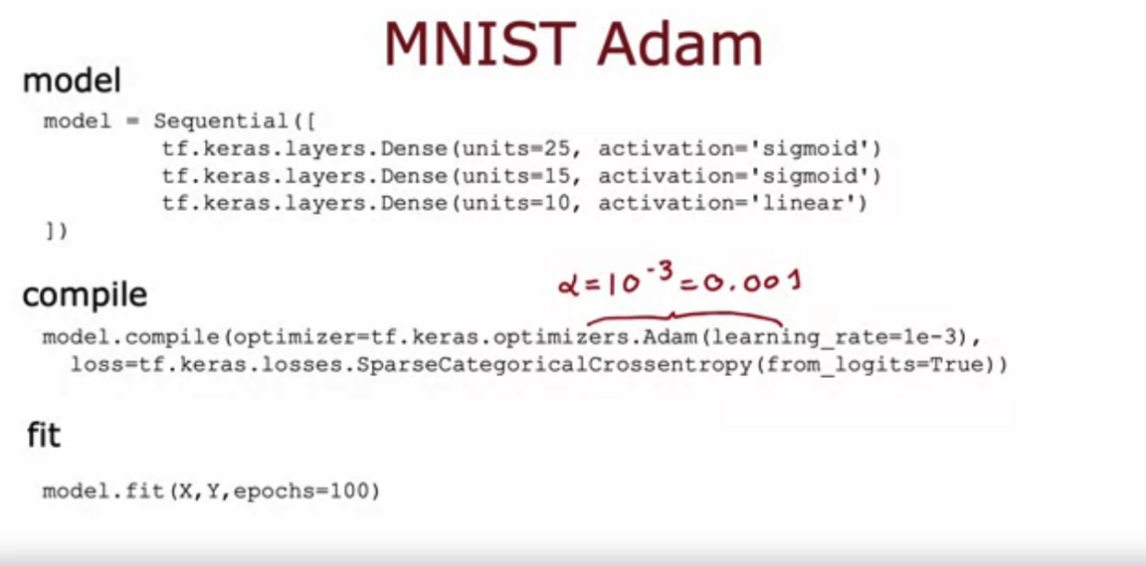
1. a 'linear' activation
2. a 'softmax' activation

Answer:A

# 第四单元

Additional Neural Network Concepts

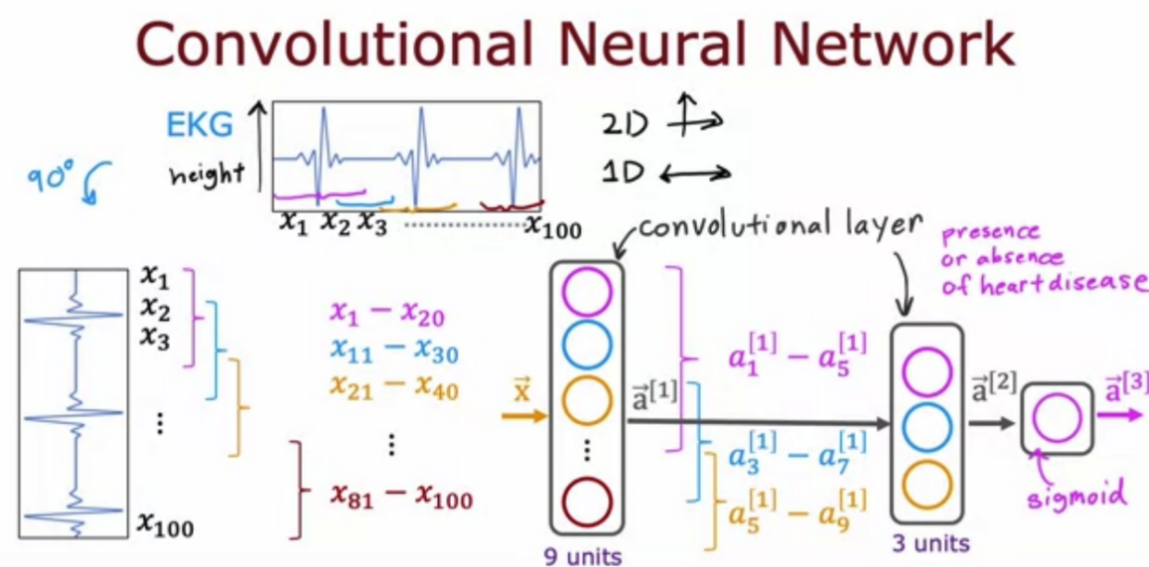
其他神经网络概念



The Adam optimizer is the recommended optimizer for finding the optimal parameters of the model. How do you use the Adam optimizer in TensorFlow?

1. The call to model.compile() uses the Adam optimizer by default
2. The Adam optimizer works only with Softmax outputs. So if a neural network has a Softmax output layer, TensorFlow will automatically pick the Adam optimizer.
3. When calling model.compile, set optimizer=tf.keras.optimizers.Adam(learning\_rate=1e-3).
4. The call to model.compile() will automatically pick the best optimizer, whether it is gradient descent, Adam or something else. So there’s no need to pick an optimizer manually.

Answer:C



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?

1. 1D layer or 2D layer (depending on the input dimension)
2. Image layer
3. A fully connected layer
4. convolutional layer

Answer:D

# 第一单元

Advice for applying machine learning

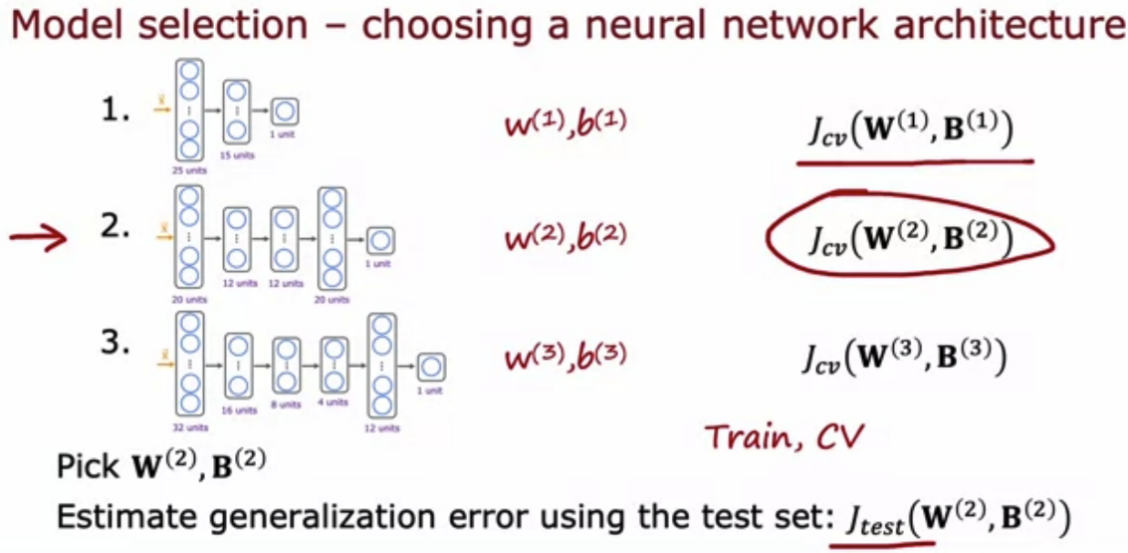
应用机器学习的建议

1. In the context of machine learning, what is a diagnostic?
2. A test that you run to gain insight into what is/isn’t working with a learning algorithm.
3. A process by which we quickly try as many different ways to improve an algorithm as possible, so as to see what works.
4. This refers to the process of measuring how well a learning algorithm does on a test set (data that the algorithm was not trained on).
5. An application of machine learning to medical applications, with the goal of diagnosing patients’ conditions.

Answer:A

1. True/False? It is always true that the better an algorithm does on the training set, the better it will do on generalizing to new data.
2. True
3. False

Answer:B



For a classification task; suppose you train three different models using three different neural network architectures. Which data do you use to evaluate the three models in order to choose the best one?

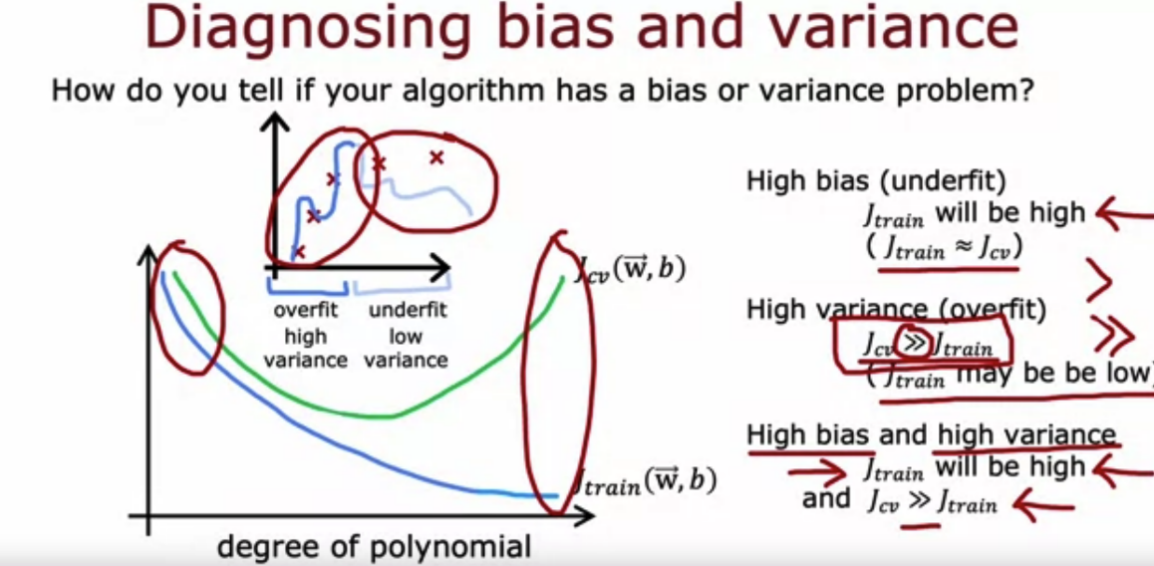
1. The test set
2. The cross validation set
3. The training set
4. All the data -- training, cross validation and test sets put together.

Answer:B

# 第二单元

Bias and variance

偏差和方差



If the model's cross validation error is much higher than the training error , this is an indication that the model has…

1. Low variance
2. high bias
3. high variance
4. Low bias

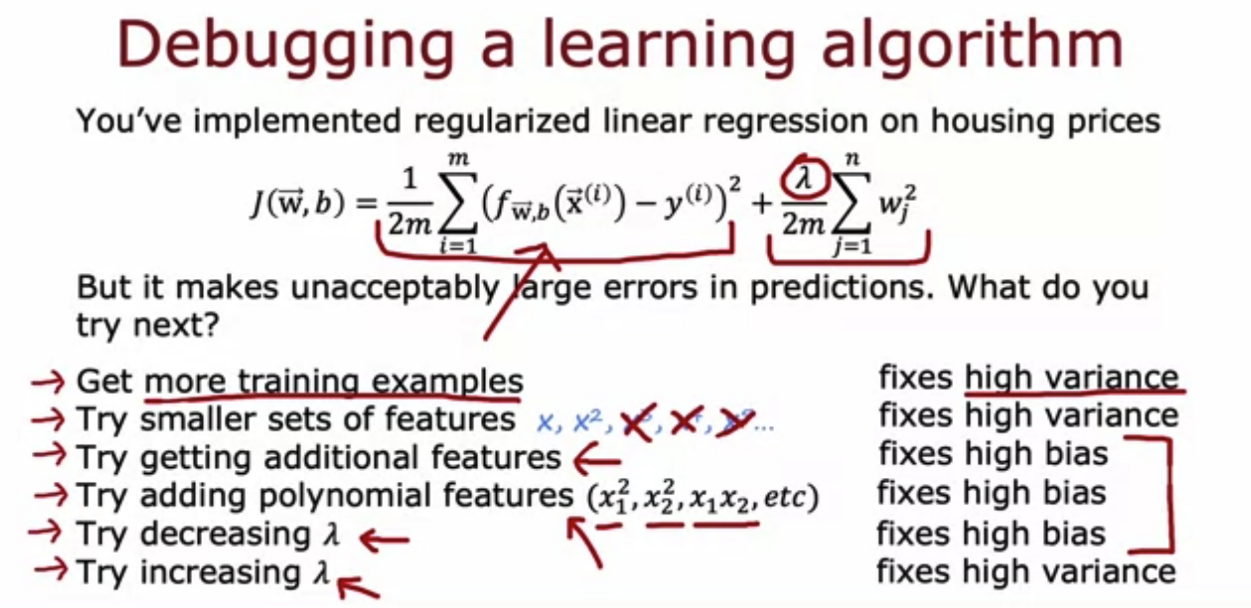
Answer:C



Which of these is the best way to determine whether your model has high bias (has underfit the training data)?

1. See if the cross validation error is high compared to the baseline level of performance
2. See if the training error is high (above 15% or so)
3. Compare the training error to the cross validation error.
4. Compare the training error to the baseline level of performance

Answer:D



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.

1. Collect additional features or add polynomial features
2. Decrease the regularization parameter λ (lambda)
3. Collect more training examples
4. Remove examples from the training set

Answer: B、C

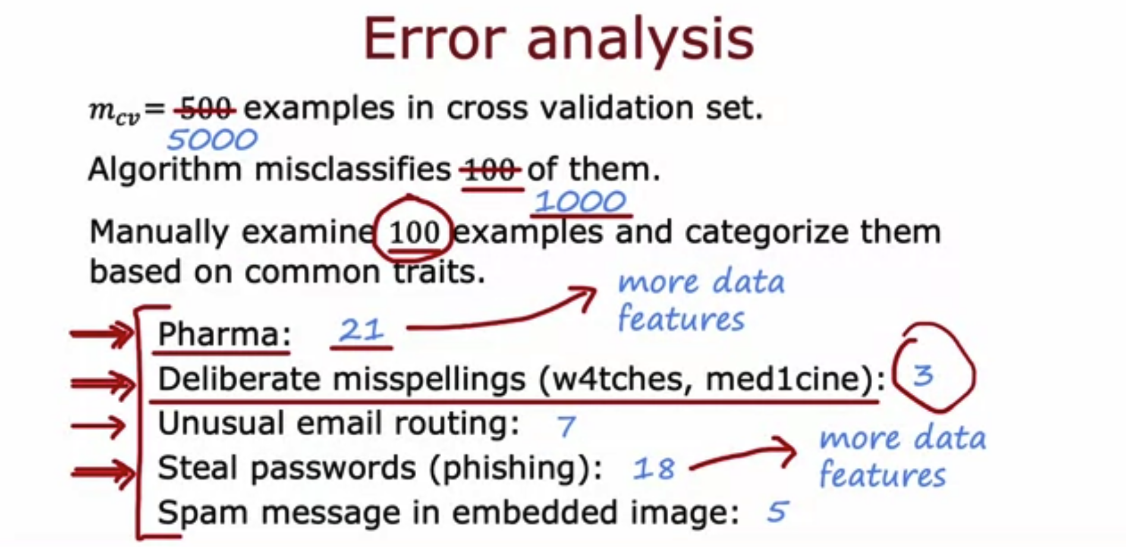
1. You find that your algorithm has a training error of 2%, and a cross validation error of 20% (much higher than the training error). Based on the conclusion you would draw about whether the algorithm has a high bias or high variance problem, which of these seem like good options for improving the algorithm’s performance? Hint: two of these are correct.
2. Collect more training data
3. Reduce the training set size
4. Increase the regularization parameter λ
5. Decrease the regularization parameter λ

Answer: A、C

# 第三单元

Machine learning development process

机器学习开发流程

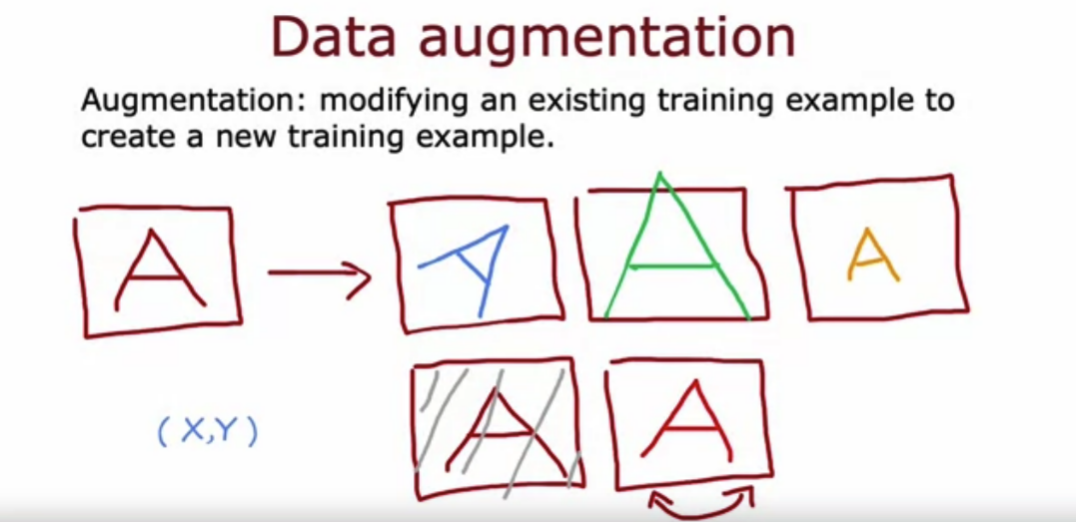


Which of these is a way to do error analysis?

1. Collecting additional training data in order to help the algorithm do better.
2. Calculating the test error
3. Calculating the training error
4. Manually examine a sample of the training examples that the model misclassified in order to identify common traits and trends.

Answer:

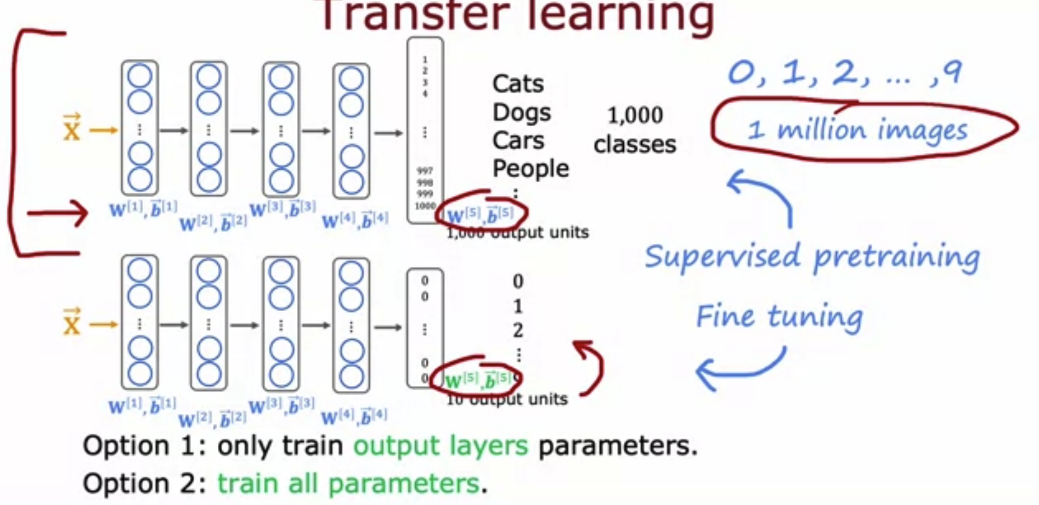
1. D



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?

1. Error analysis
2. Machine learning diagnostic
3. Data augmentation
4. Bias/variance analysis

Answer:C



What are two possible ways to perform transfer learning? Hint: two of the four choices are correct.

1. You can choose to train just the output layers' parameters and leave the other parameters of the model fixed.
2. Download a pre-trained model and use it for prediction without modifying or re-training it.
3. Given a dataset, pre-train and then further fine tune a neural network on the same dataset.
4. You can choose to train all parameters of the model, including the output layers, as well as the earlier layers.

Answer:D