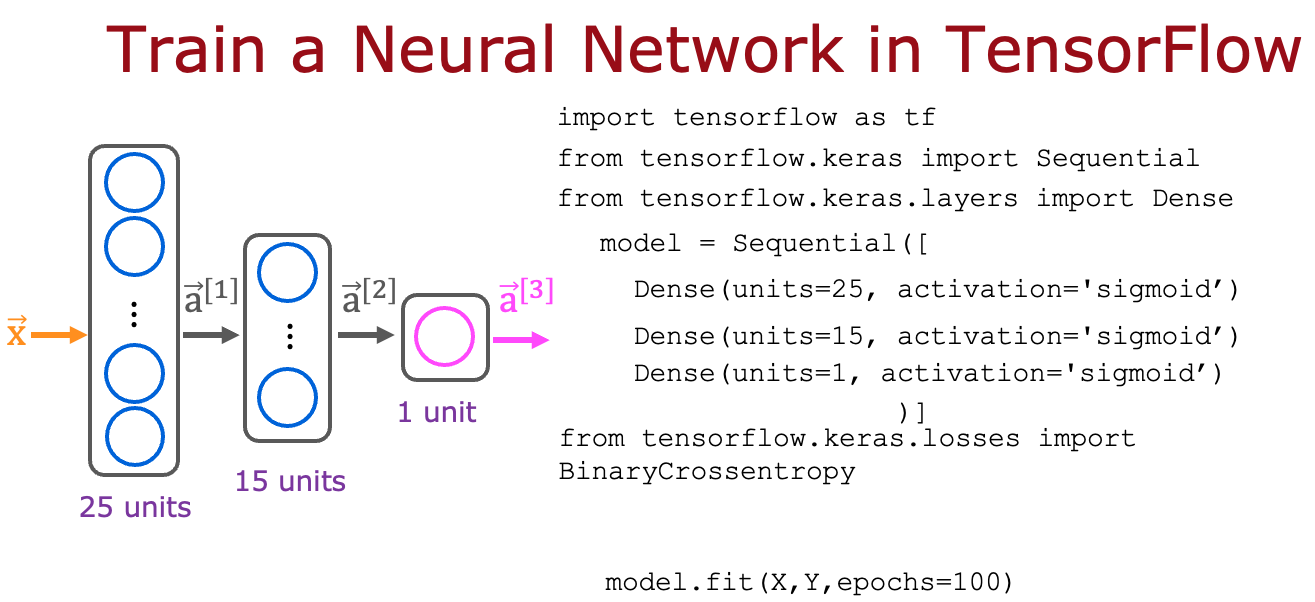
# 第一单元

Neural Network Training

神经网络的训练



Here is some code that you saw in the lecture:

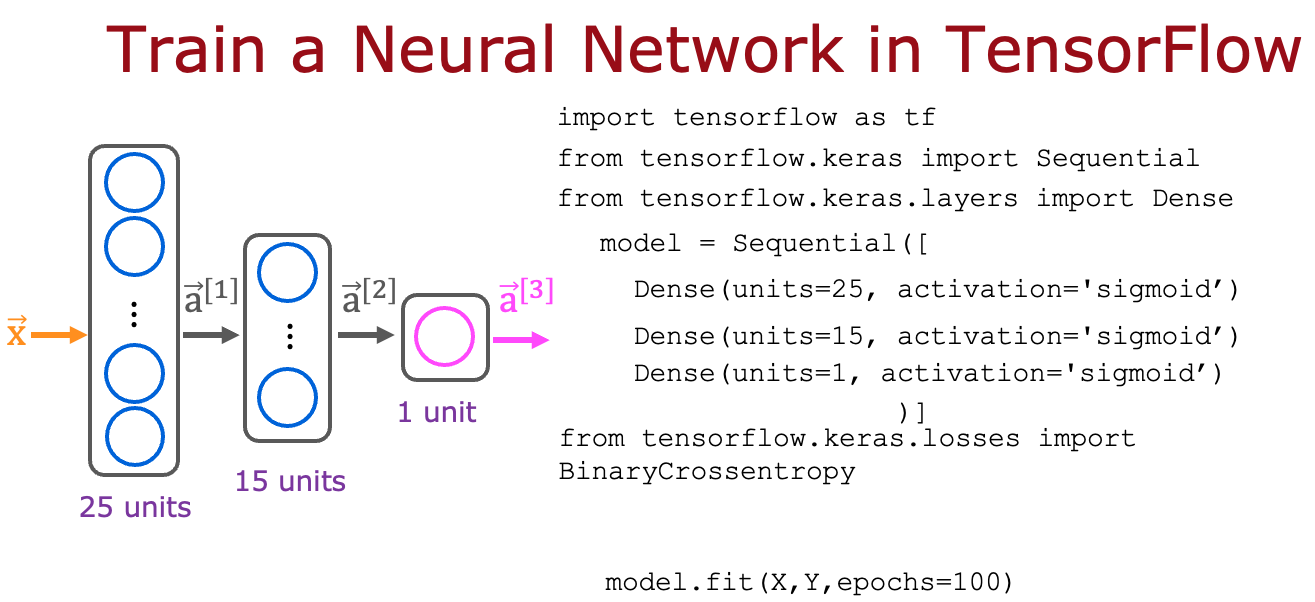
```

model.compile(loss=BinaryCrossentropy())

```

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

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)



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)

```

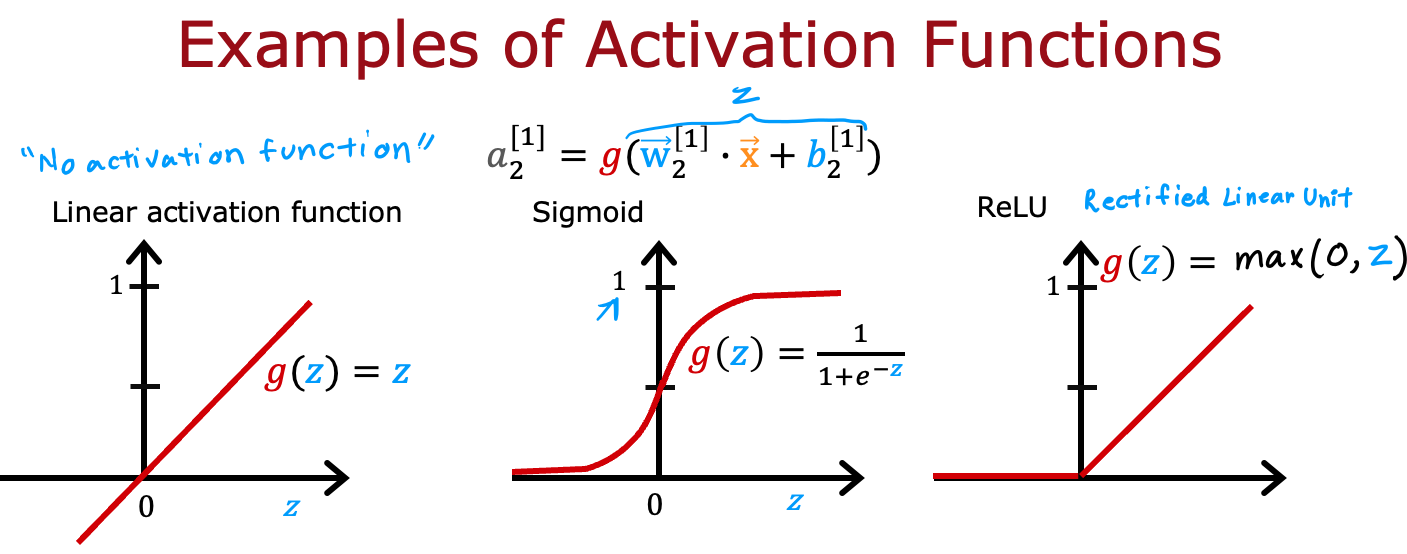
Which line of code updates the network parameters in order to reduce the cost? D

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)

# 第二单元

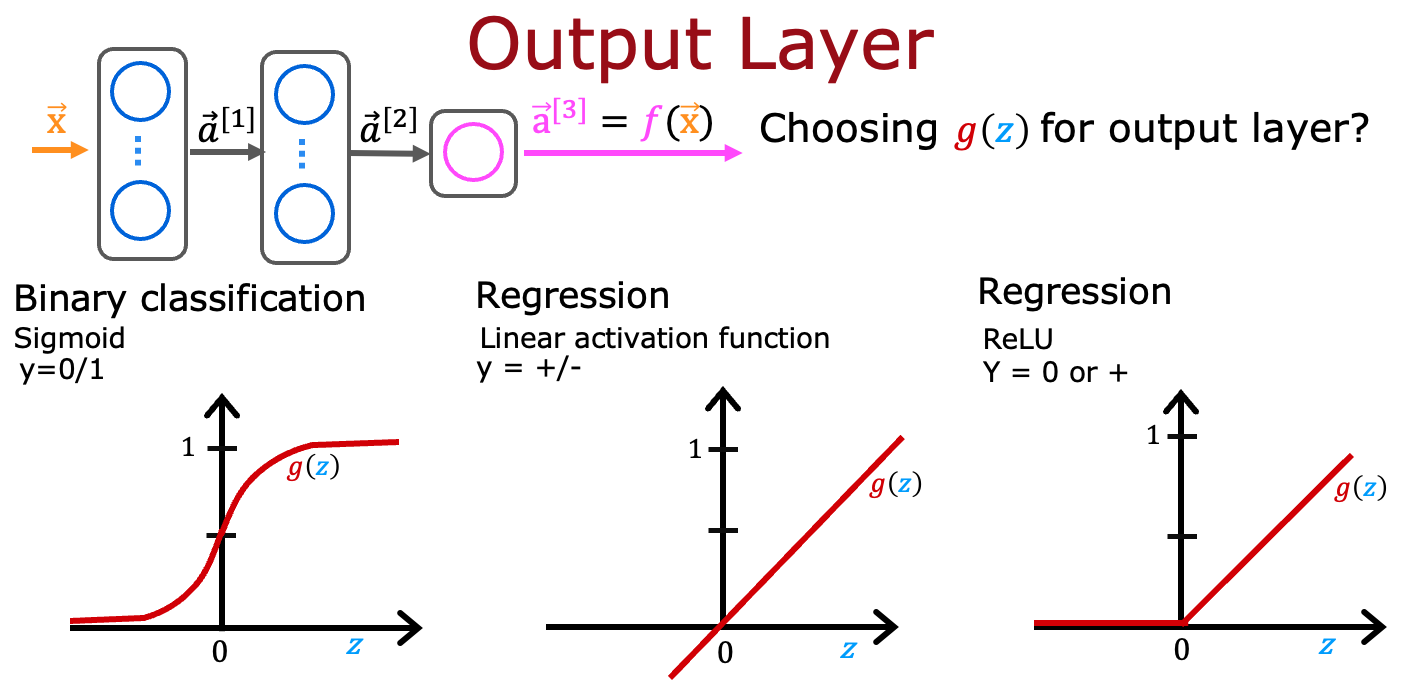
Activation Functions

激活函数



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

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



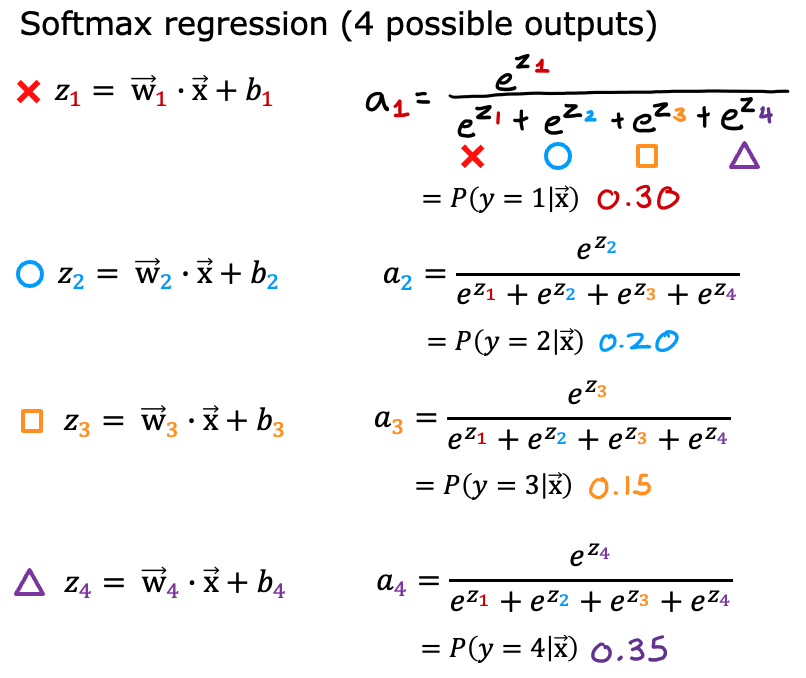
For the task of predicting housing prices, which activation functions could you choose for the output layer? Choose the 2 options that apply. B

1. Sigmoid
2. linear
3. ReLU
4. 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. A
5. True
6. False

# 第三单元

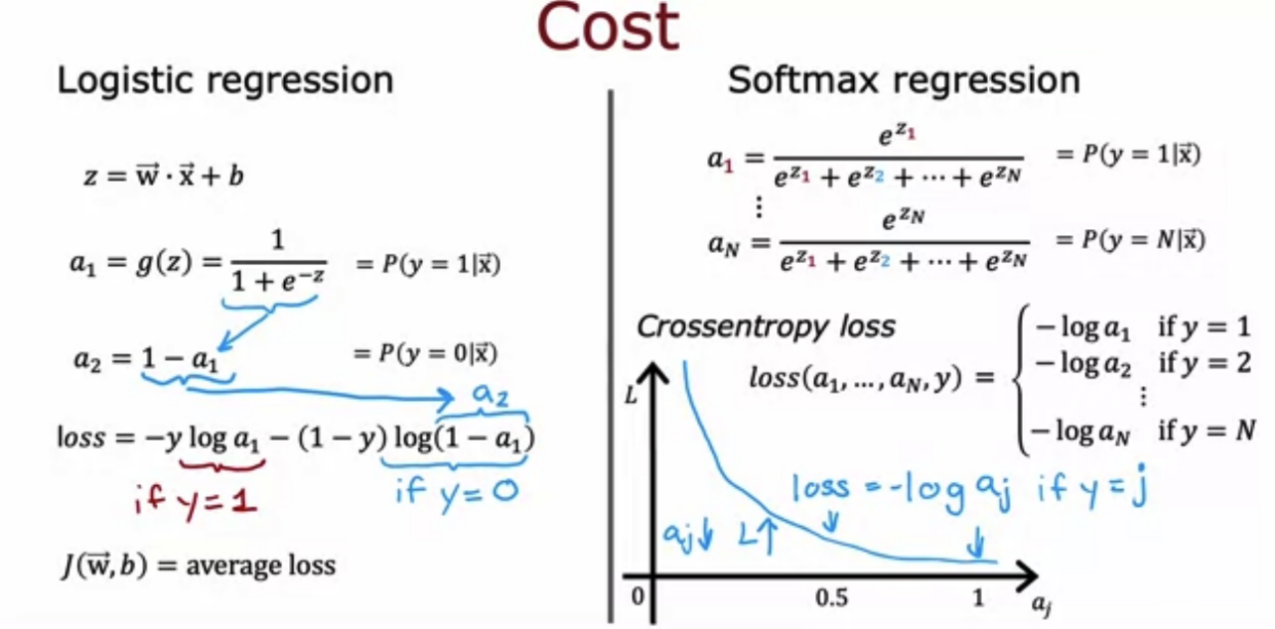
Multiclass Classification

多类别分类

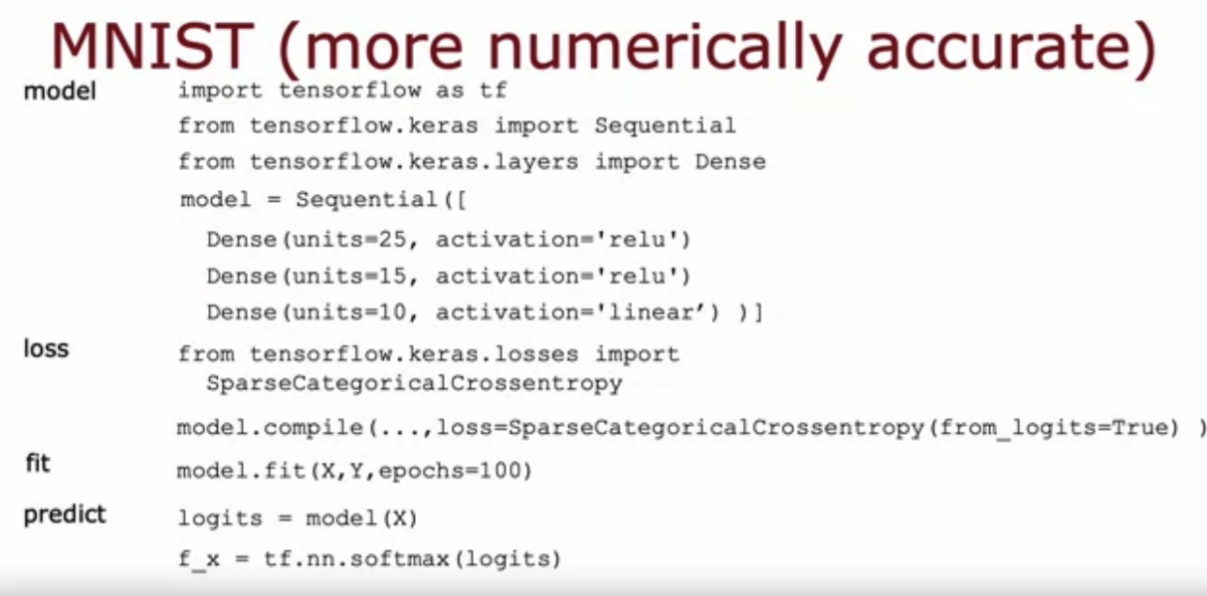


For a multiclass classification task that has 4 possible outputs, the sum of all the activations adds up to 1. For a multiclass classification task that has 3 possible outputs, the sum of all the activations should add up to …. D

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



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.] D



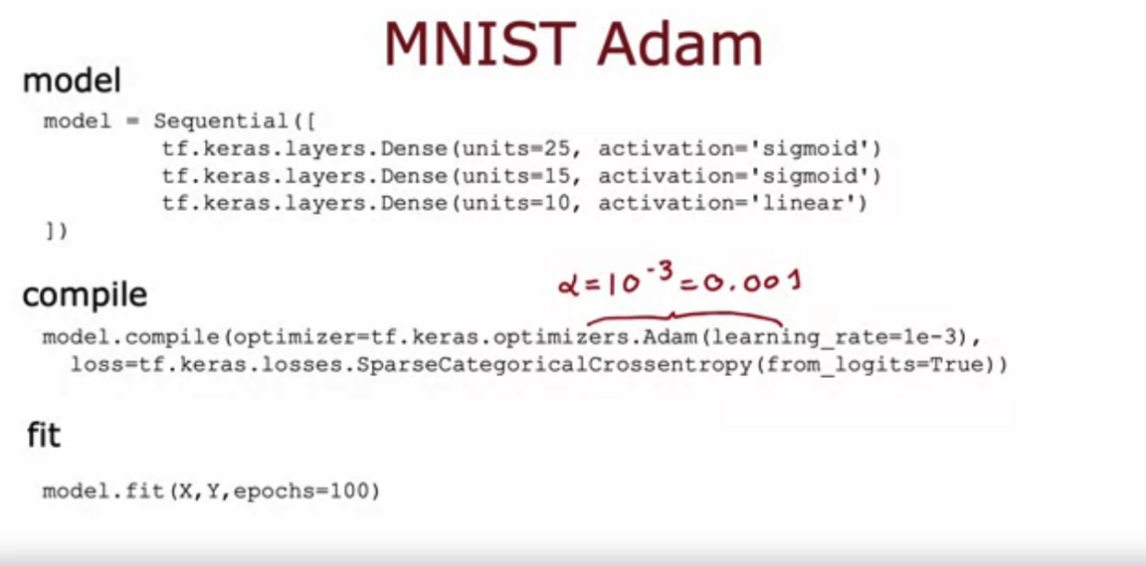
For multiclass classification, the recommended way to implement softmax regression is to set from\_logits=True in the loss function, and also to define the model's output layer with… D

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

# 第四单元

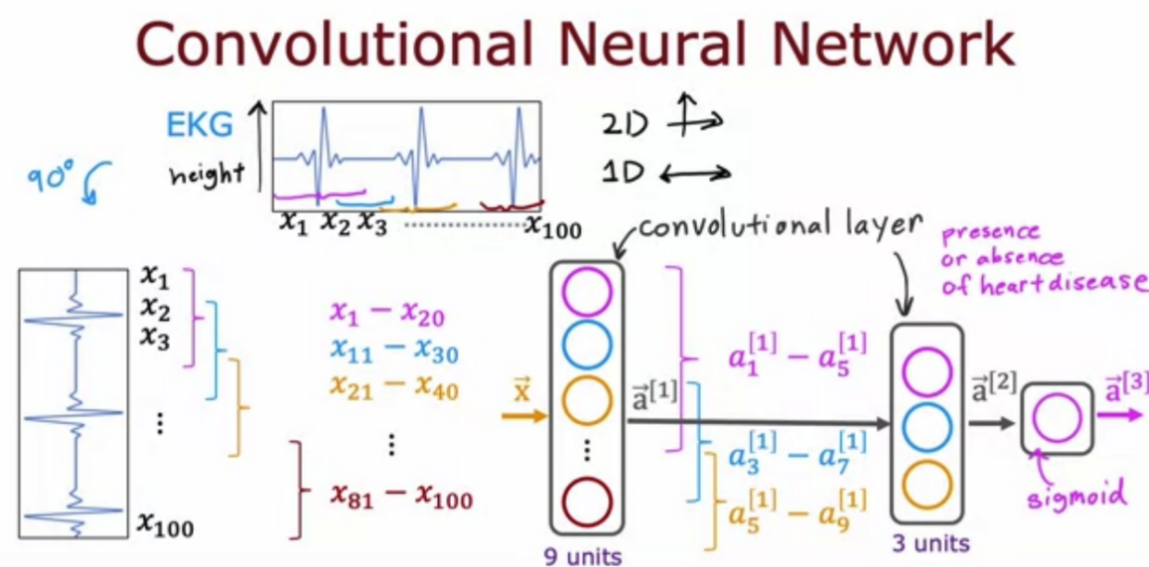
Additional Neural Network Concepts

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? C

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.



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? D

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