**What is the sigmoid function? Why is it used in logistic regression?**

The sigmoid function is:  
  
 σ(z) = 1 / (1 + e^(-z))  
  
It compresses any input into the range (0, 1), making it best for expressing probabilities. In logistic regression, it compresses the linear output z = wx + b into a probability.

**What is binary cross-entropy loss?**

Binary cross-entropy measures the error between predicted probability and actual binary label:  
  
 Loss = -[y \* log(p) + (1 - y) \* log(1 - p)]  
  
Where y is the actual label and p is the value we got through sigmoid function

**If the predicted probability is 0.8 and the true label is 1, compute the binary cross-entropy loss.**

Given that y = 1 and p = 0.8

We know that Loss = -[y \* log(p) + (1 - y) \* log(1 - p)]

Therefore Loss = -[1 \* log(0.8) + (1 - 1) \* log(1 – 0.8)]  
 = -[1 \* log(0.8) + (0) \* log(0.2)]  
 = -log(0.8)

**What does it mean when the output of logistic regression is 0.5?**

It shows that the model is uncertain because it has assigned each class a 50% chance. The model would classify the input as class 1 if p > 0.5 else class 0.

**Derive the gradient of the binary cross-entropy loss with respect to the weight w.**

Prediction: p = Sigmoid(z)

z = wx + b  
  
Loss: L = -[y \* log(p) + (1 - y) \* log(1 - p)]  
  
Using the chain rule: dL/dw = dL/dp \* dp/dz \* dz/dw

dL/dp = -y/p + (1-y)/(1-p)

dp/dz = p(1-p)

dz/dw = x

Therefore dL/dw = {-y/p + (1-y)/(1-p)}\*{ p(1-p)}\*x  
 = (p-y)x

Similarly, dL/db = (p-y)

**How does gradient descent work in logistic regression?**

Gradient descent updates the weight and bias using the gradients of the loss:  
  
 w := w - a \* dL/dw  
 b := b - a \* dL/db  
  
Where a is the learning rate. This process is repeated over many epochs to minimize the loss.

**What are epochs in machine learning?**

An epoch is one complete cycle through the training dataset. During each epoch, the model updates its parameters using gradient descent to improve accuracy.

**Explain the purpose of the bias term in logistic regression.**

The bias term b allows the model to make predictions even when input features like w and x are zero. It shifts the decision boundary left or right to better fit the data.

**What’s the decision boundary in logistic regression and how is it derived?**

The decision boundary is where the predicted probability is 0.5  
  
 1 / (1 + e^-(wx + b)) = 0.5 ⇒ wx + b = 0  
  
So the decision boundary is the line wx + b = 0. Inputs on one side are classified as 0 and on the other side as 1.

**If logistic regression outputs a probability of 0.3 for a sample and the actual class is 1, is this a good prediction? Why or why not?**

No. A prediction of 0.3 for a true class of 1 means the model was highly inaccurate. This would result in a high binary cross-entropy loss. The model needs to adjust the weights to predict better.