Foundations Of Neural Networks and Deep Learning

Day-5

SME: Gagan P

Contact: gaganp000999@gmail.com

recap:

1. What is the main objective of Gradient Descent?

A. To increase the loss function
B. To minimize the loss function by adjusting parameters
C. To maximize accuracy directly
D. To randomly update weights and bias

recap:

1. What is the main objective of Gradient Descent?

A. To increase the loss function
B. To minimize the loss function by adjusting parameters
C. To maximize accuracy directly
D. To randomly update weights and bias

2. In Gradient Descent, the “gradient” refers to:

A. The data distribution
B. The derivative of the loss function
C. The slope of the prediction line
D. The bias term

2. In Gradient Descent, the “gradient” refers to:

A. The data distribution

B. The derivative of the loss function

C. The slope of the prediction line

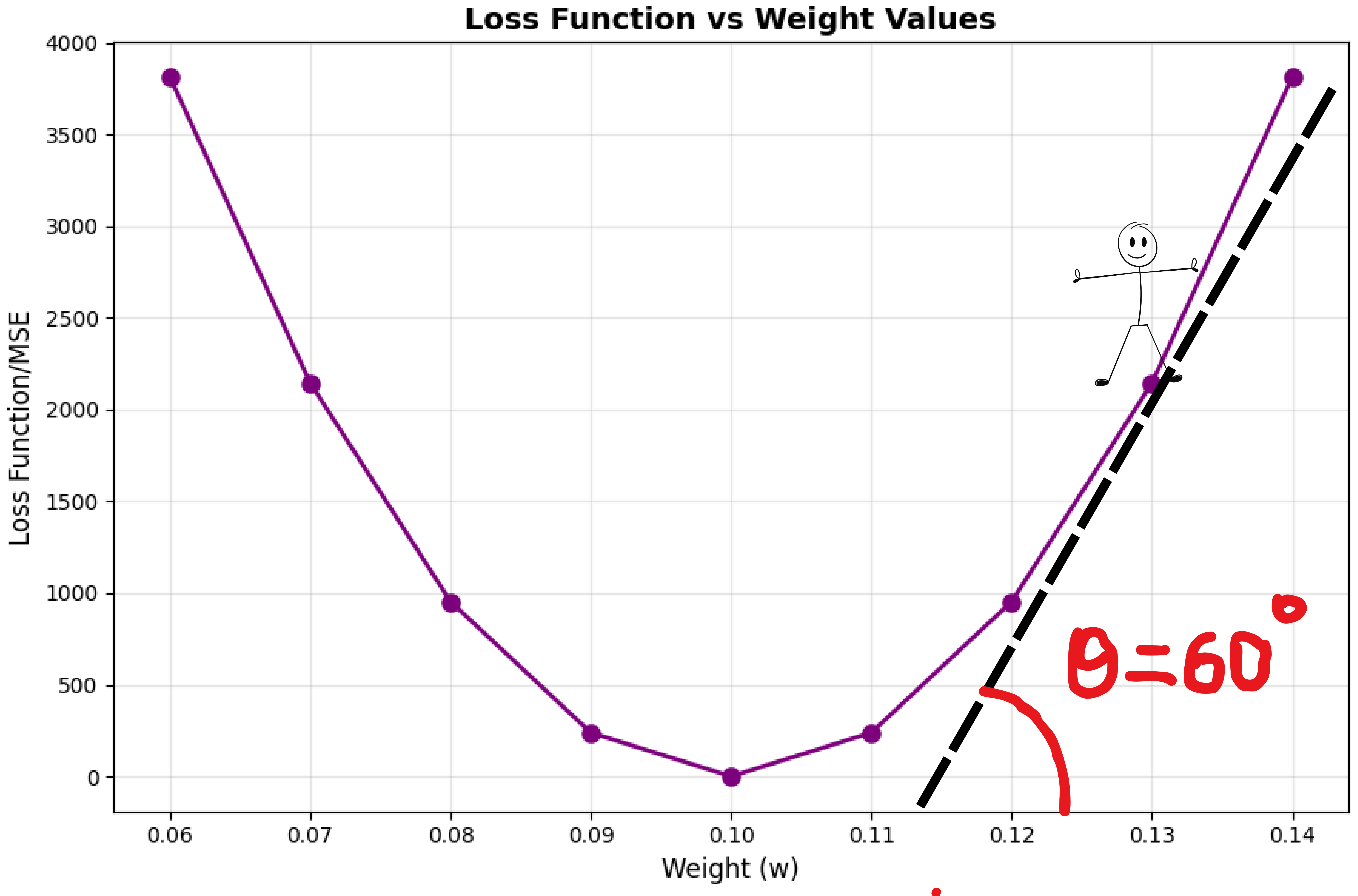
D. The bias term

If $\partial L/\partial w$ (gradient w.r.t weight) is positive, which direction should we move?

A. Increase the weight
B. Decrease the weight
C. Keep the weight constant
D. Reverse the learning rate

If ∂L/∂w (gradient w.r.t weight) is positive, which direction should we move?

A. Increase the weight
B. Decrease the weight
C. Keep the weight constant
D. Reverse the learning rate

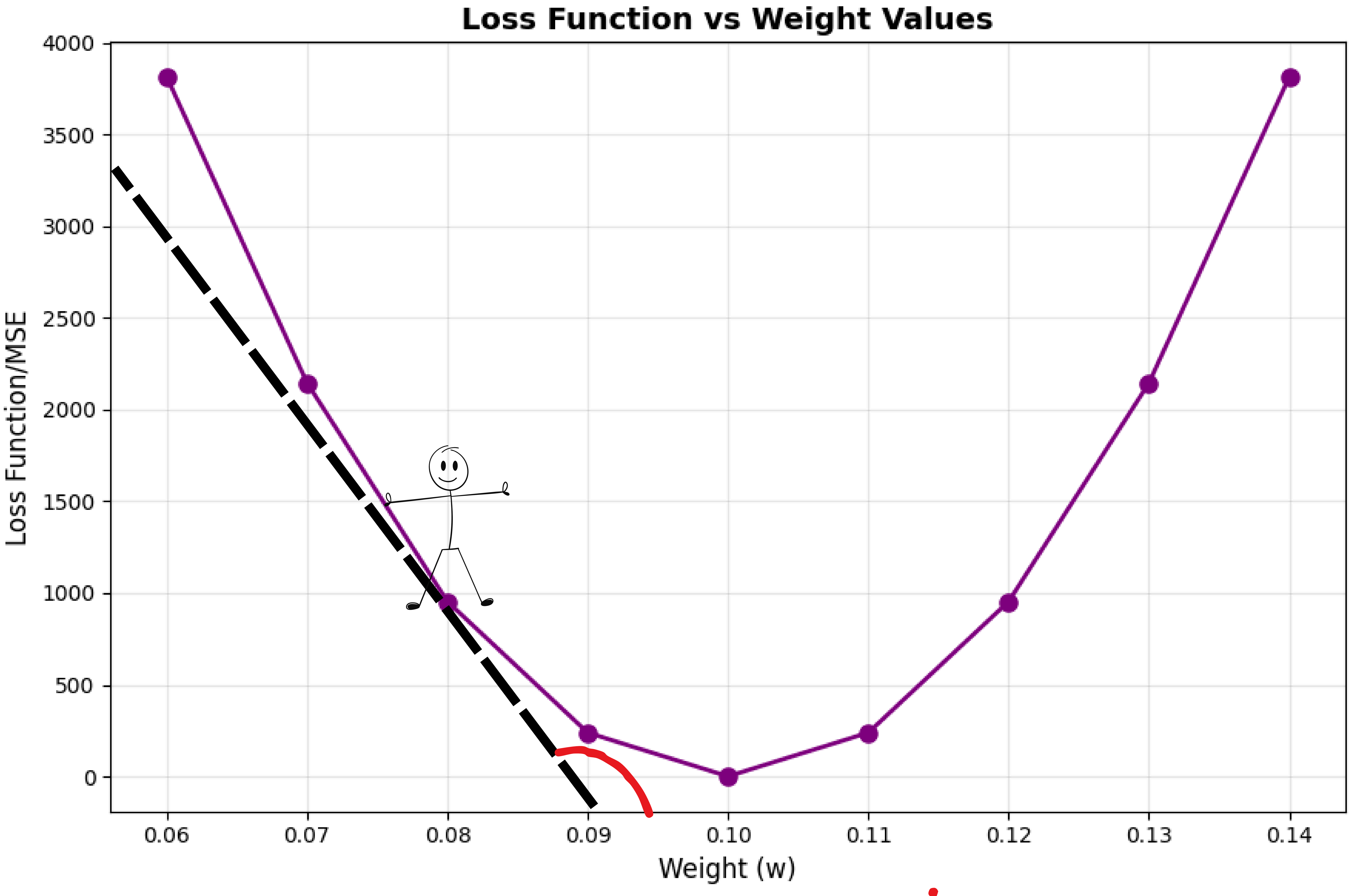

Loss Function vs Weight Values
Loss Function/MSE
Weight (w)
0
500
1000
1500
2000
2500
3000
3500
4000
0.06
0.07
0.08
0.09
0.10
0.11
0.12
0.13
0.14
θ=60°

What does a negative gradient indicate in the context of loss vs. weight graph?

A. Loss increases when weight increases
B. Loss decreases when weight increases
C. No relation between weight and loss
D. Model is unstable

What does a negative gradient indicate in the context of loss vs. weight graph?

A. Loss increases when weight increases
B. Loss decreases when weight increases
C. No relation between weight and loss
D. Model is unstable



Loss Function vs Weight Values
Loss Function/MSE
4000
3500
3000
2500
2000
1500
1000
500
0
0.06
0.07
0.08
0.09
0.10
0.11
0.12
0.13
0.14
Weight (w)

The parameter α (alpha) in gradient descent is known as:

A. Regularization term
B. Bias factor
C. Learning rate(Step Size)
D. Step count

The parameter α (alpha) in gradient descent is known as:

A. Regularization term
B. Bias factor
C. Learning rate(Step Size)
D. Step count

What happens if the learning rate is too high?

A. Model converges faster
B. Model might overshoot and never converge
C. Gradient becomes zero immediately
D. Model accuracy increases

What happens if the learning rate is too high?

A. Model converges faster
B. Model might overshoot and never converge
C. Gradient becomes zero immediately
D. Model accuracy increases

Which of the following best represents the weight update rule in Gradient Descent?

A. $w = w + \alpha \times \partial L/\partial w$
B. $w = w - \alpha \times \partial L/\partial w$
C. $w = \alpha \times \partial L/\partial w$
D. $w = w \times \alpha \times L$

Which of the following best represents the weight update rule in Gradient Descent?

A. w = w + α × ∂L/∂w

B. w = w - α × ∂L/∂w

C. w = α × ∂L/∂w

D. w = w × α × L

$$w := w - \alpha \frac{dL}{dw}$$

In Gradient Descent, what does the slope of the loss curve represent?

A. The rate of change of weights
B. The rate at which loss changes with respect to weights
C. The curvature of the data
D. The learning rate magnitude

In Gradient Descent, what does the slope of the loss curve represent?

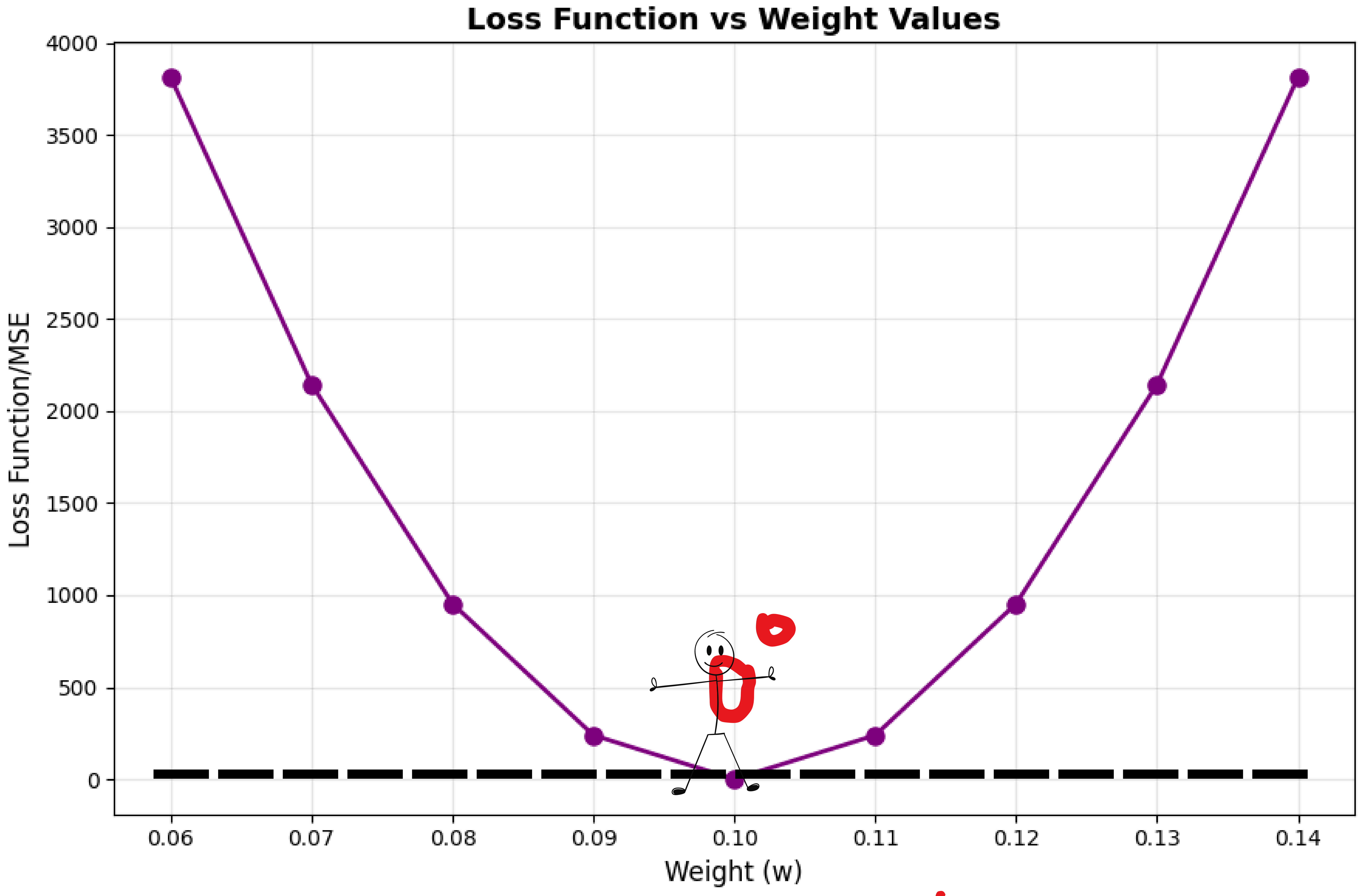
A. The rate of change of weights
B. The rate at which loss changes with respect to weights
C. The curvature of the data
D. The learning rate magnitude

When the derivative (gradient) of the loss function becomes zero, it means:

A. The model is overfitting
B. The model has reached a minimum point
C. The learning rate is too small
D. The cost function has diverged

When the derivative (gradient) of the loss function becomes zero, it means:

A. The model is overfitting
B. The model has reached a minimum point
C. The learning rate is too small
D. The cost function has diverged



Loss Function vs Weight Values
4000
3500
3000
2500
2000
1500
1000
500
0
Loss Function/MSE
0.06
0.07
0.08
0.09
0.10
0.11
0.12
0.13
0.14
Weight (w)

what we learnt last class

$$L(w,b) = \frac{1}{n}\sum_{i=1}^{n}(y_i - (wx_i + b))^2$$

what we learnt last class

$$w := w - \alpha \frac{\partial L}{\partial w}$$

$$b := b - \alpha \frac{\partial L}{\partial b}$$

$$\frac{\partial L}{\partial w} = \frac{-2}{n} \sum_{i=1}^{n} x_i(y_i - (wx_i + b))$$

$$\frac{\partial L}{\partial b} = \frac{-2}{n} \sum_{i=1}^{n} (y_i - (wx_i + b))$$

what we learnt last class

```
for epoch in range(epochs):
    y_pred = w * X + b
    dw = (-2/n) * np.sum(X * (y - y_pred))
    db = (-2/n) * np.sum(y - y_pred)
    w = w - alpha * dw
    b = b - alpha * db
```

day 5 - training a linear regression model on california housing dataset

hands on session