Multi class Logistic Regression

$\hat{y} = \sigma(z)$: prob

$z = w^T x$

prob > thresh ⟶ class A

⟶ class B

Mohd Ayyoob
(M.Tech CSA, IISc Banglore)

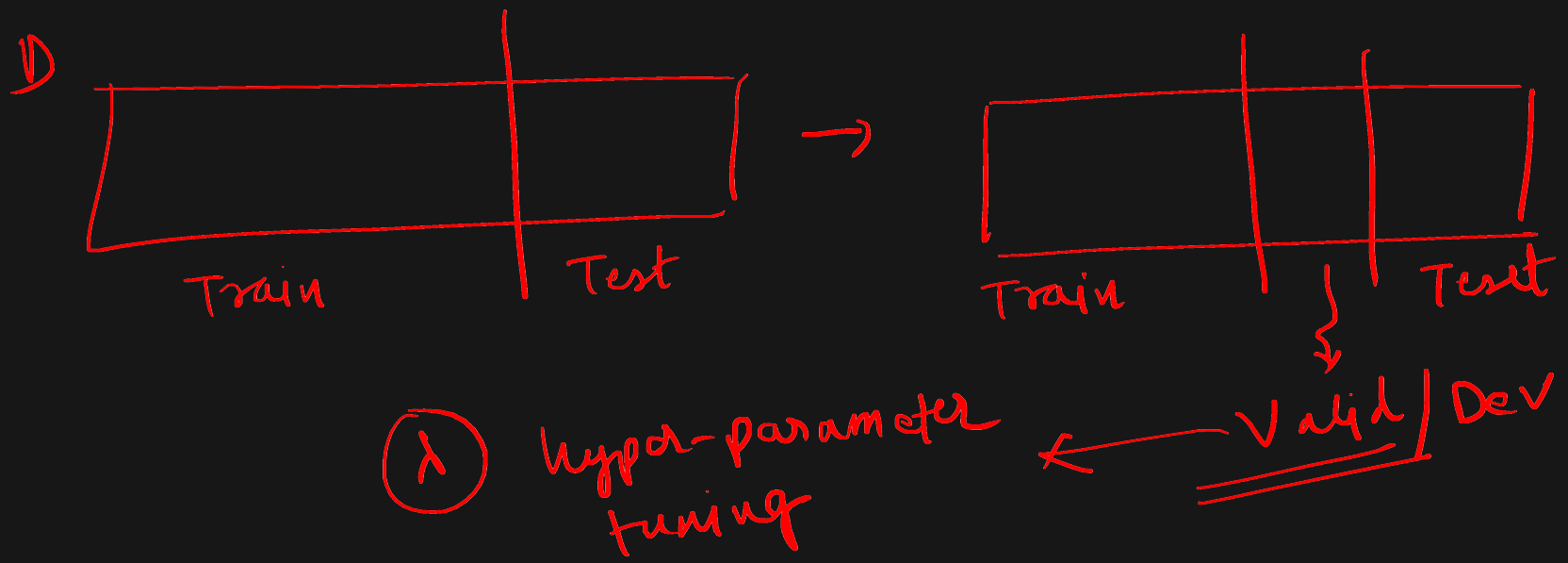
Q1. For logistic regression, the amount of L2 regularization λ should be chosen to maximize accuracy on the test set.

Should be completely unseen

Hyper-parameter

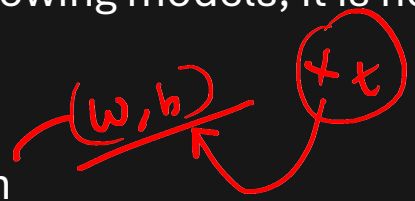
True

~~False~~



MSQ

Q2. For which of the following models, it is necessary to keep the training data stored in memory



a. Linear Regression
b. k-Nearest Neighbors
c. [redacted] Logistic Regression
d. None

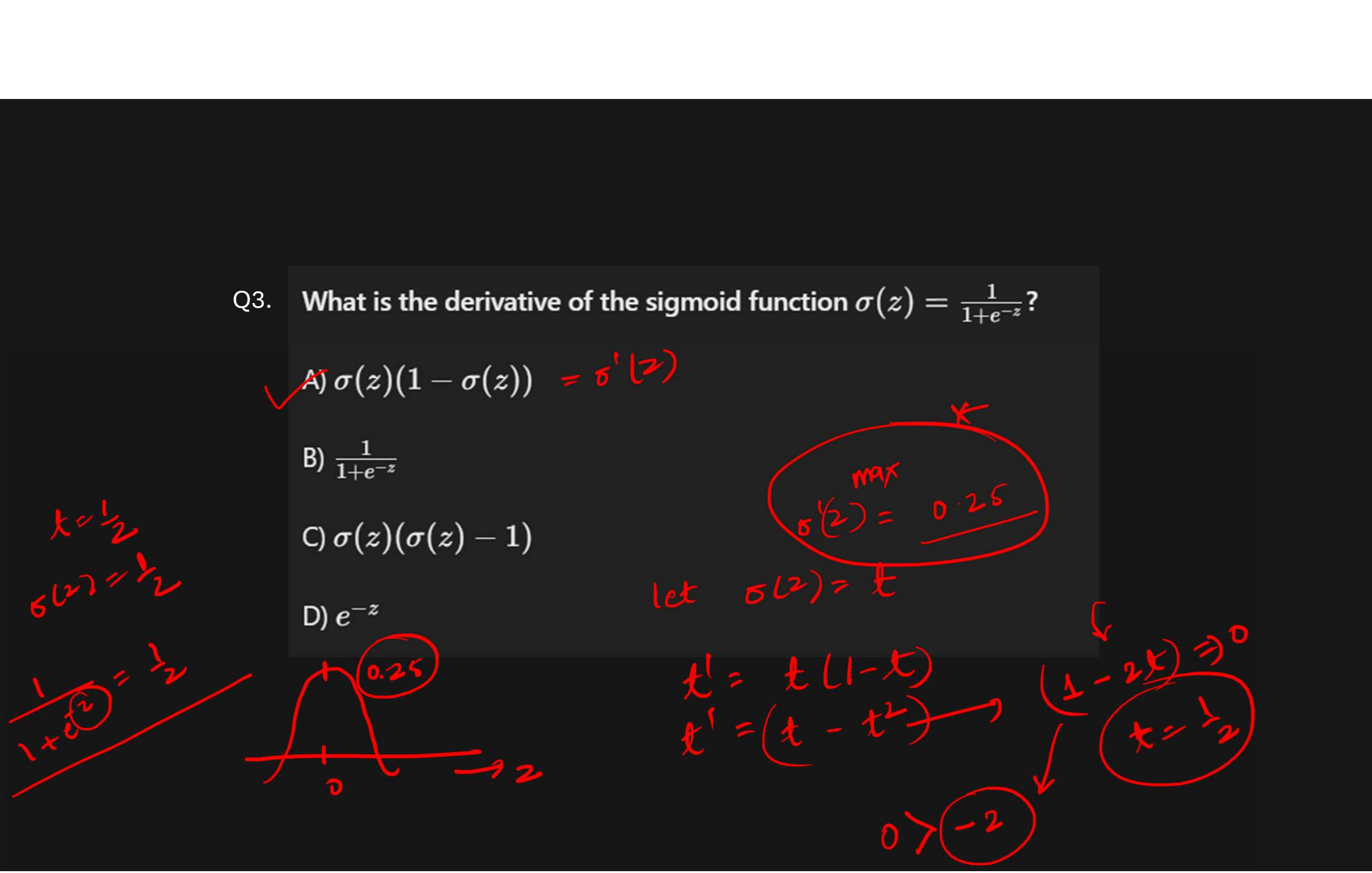
Q3. **What is the derivative of the sigmoid function $\sigma(z) = \frac{1}{1+e^{-z}}$?**

A) $\sigma(z)(1-\sigma(z))$

B) $\frac{1}{1+e^{-z}}$

C) $\sigma(z)(\sigma(z)-1)$

D) e^{-z}



Q4) Given $x_t = [1, -2]$ and three different models say, for each model of logistic regression

$w_1 = [1, 2]$, $w_2 = [1, -2]$ $w_3 = [-1, 2]$

The order of prob assigned by respective model would be

→ do you understand what is o/p in case of Log Reg model

a. m1 > m2 > m3

b. m2 > m1 > m3

c. m3 > m1 > m2

d. The model output should not be compared as they are labels in case of logistic regression

z : logit

$\sigma(\vec{w}\vec{x}) = \text{prob} > \text{thr}$

x [1, -2] w_1 = [1, 2]

$x @ w_1 = (1 \times 1) + (-2 \times 2)$

$= 1 + (-4)$

$x @ w_1 = -3$

Binary cross-entropy

$$-\left(\sum_{i=1}^{N} y_i \log(h_i) + (1-y_i)\log(1-h_i)\right)$$

$y_i \in \{0, 1\}$

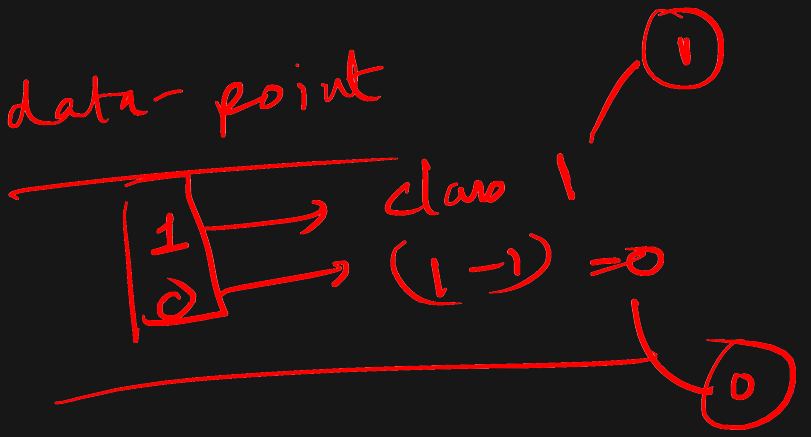
y_i (true label)

$\hat{y}_i$

$h_i \in [0, 1]$

N : # eg/ samp/ data-point

h_i = predic^n of model = $\sigma(w^T x)$: $\hat{y}_i$



Multi-class - cross entropy

- CATEGORICAL

$$\sum_{i=1}^{N} \sum_{j=1}^{C=4} \left(y_j^i * -\log\left(h_j^i\right) \right)$$

y [0, 0, 1, 0]

x = [- - - -]

sigmoid(xw)

[(-) - - -]

[0, 0, 0, 0]

(0, 1)

wtd -ve log prob

N: # eg

C: # classes

actual label

neg log prob

$C \geq 2$

true-labels (y): [1 0]

class A, B, C

x¹

x²

x³

A

B

C

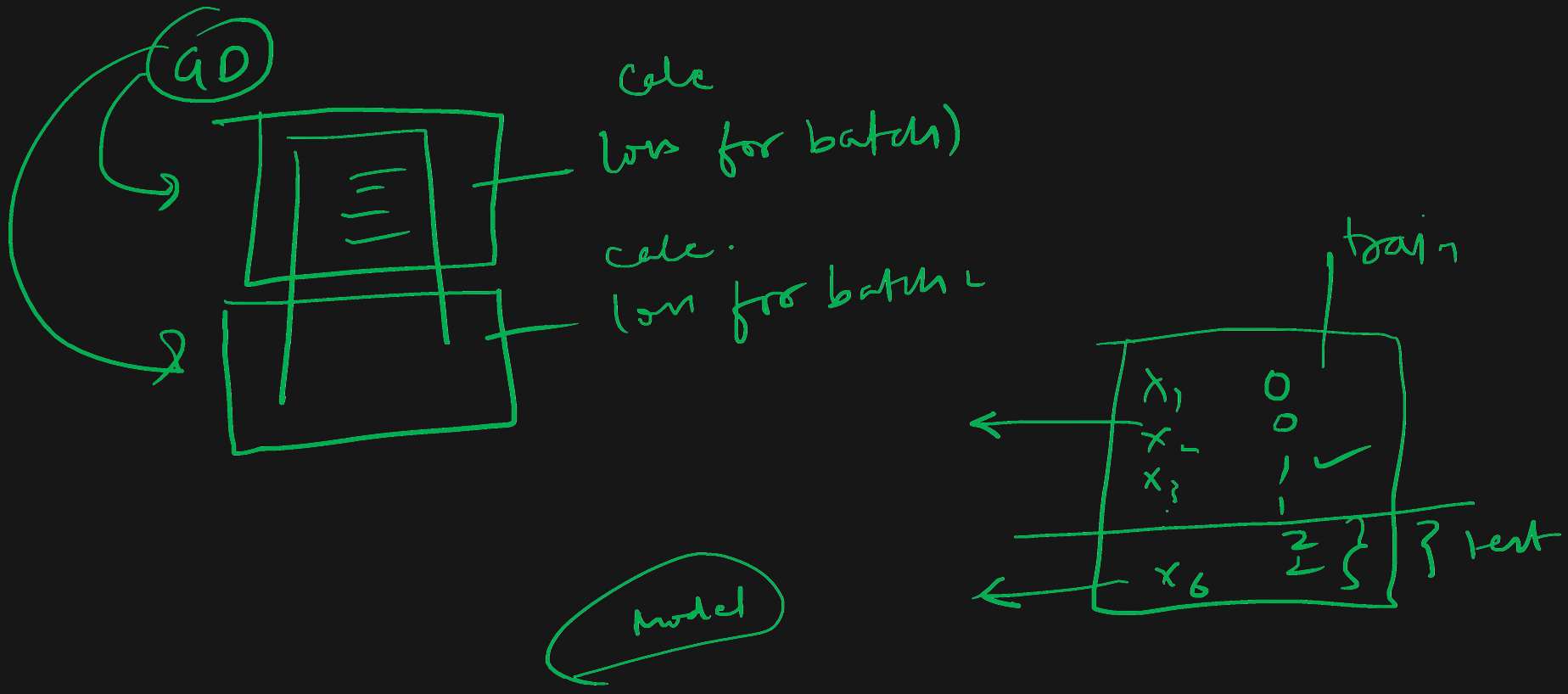
This assumes one-hot encoded labels

one-hot encoding

y¹ 1 0 0

y² 0 1 0

y³ 0 0 1


GD
calc
loss for batch)
calc:
loss for batch
model
train
test