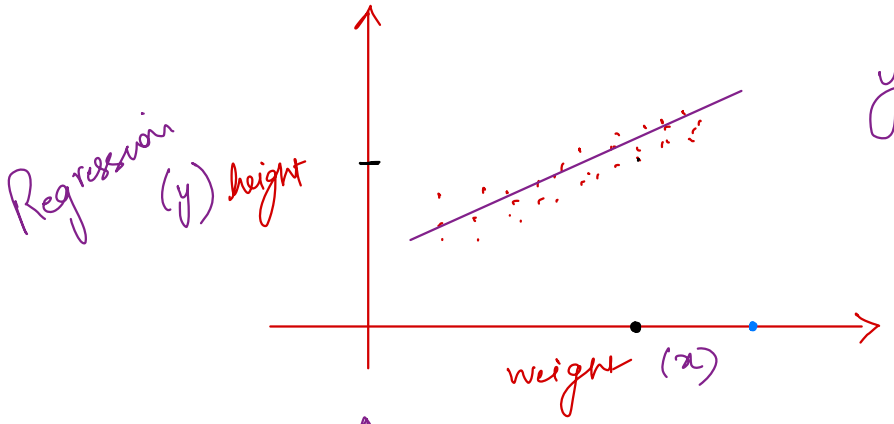
Linear Regression

↳ predicting continuous / numerical values



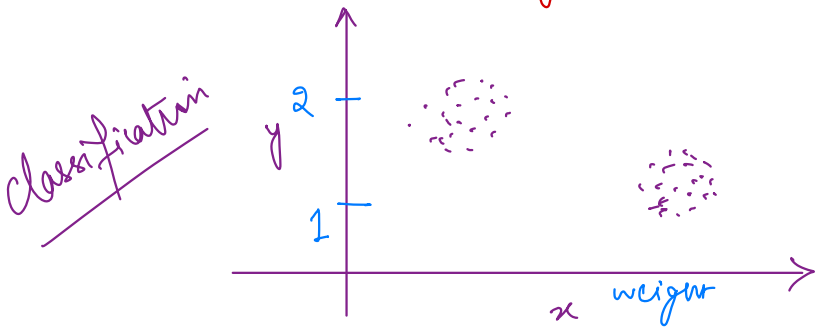
$y = mx + c$

m = w (weight)

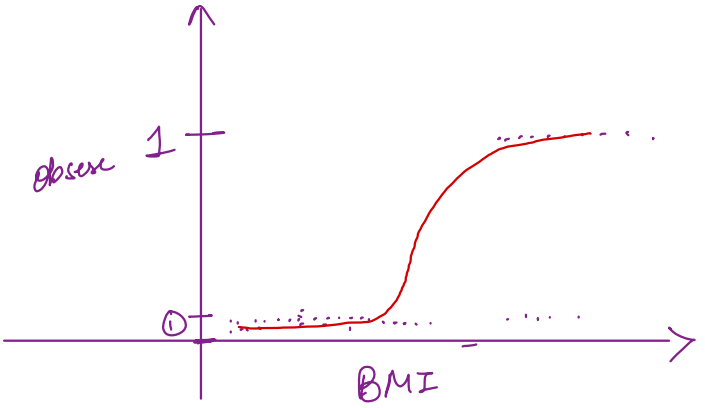
c = b (bias)

$y = wx + b$

parameters



Which of the following techniques is used to address overfitting in linear regression?
a) Ridge regression.
b) Normalization.
c) Principal Component Analysis (PCA).
d) Variance Inflation Factor (VIF).



$$y = wx + b \Rightarrow (y)$$

$$\sigma(y) = \frac{1}{1+e^{-y}} = 0 \le \sigma(y) \le 1$$

$e \approx 2.71$

$y = -\infty \qquad \sigma(y) \to 0$

$y = \infty \qquad \sigma(y) \to 1$

$y = 0 \qquad \sigma(y) = 0.5$

→ Do we want best fit line?

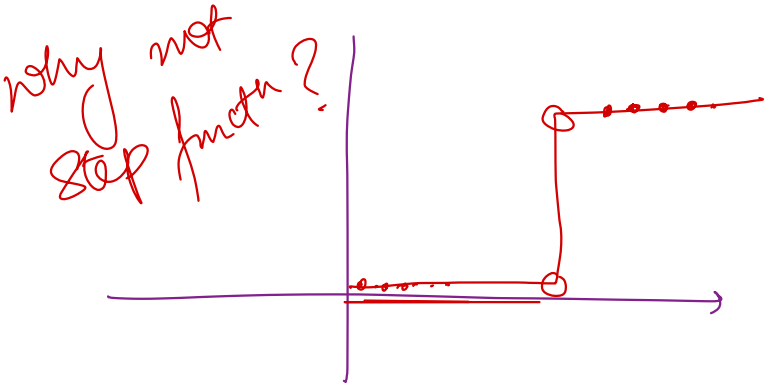
NO

→ Do we want best line for separation.

Yes

→ Supervised algorithm

→ binary classification



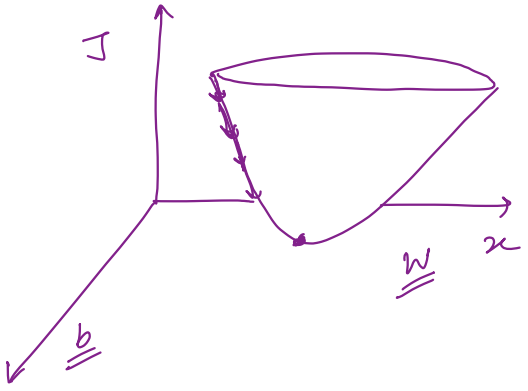
What happens when the input to the sigmoid function is a very large negative value?

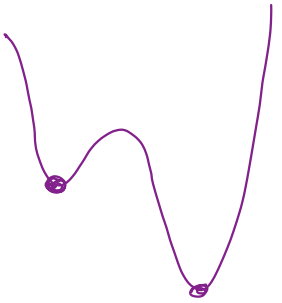
A) The output becomes negative
B) The output approaches 0
C) The output approaches 1
D) The output becomes undefined.

A company is building a credit scoring model to predict risk of default.
Which function can be used to map the model's output to a probability between 0 and 1?
A) Sigmoid function.
B) Linear function.
C) Step function.
D) Exponential function.
Correct Answer: A) Sigmoid function.

Probability & Likelihood ? (HW)

$= p^{y}(1-p)^{(1-y)}$





Log - likelihood

$$= y \ln(p) + (1-y) \ln(1-p)$$

Negative log likelihood

$$-\left(\sum_{i=1}^{m} y^{(i)} \log(p^i) + (1-y^i) \log(1-p^i) \right)$$

$$-y^{(i)} \log(\hat{y}^i) - (1-y^i) \log(1-\hat{y}^i)$$

MSE + Sigmoid → Non Convex

log loss + Sigmoid = Convex