# CIS 678 – Machine Learning

Predictive modeling: Classification

# Regression vs Classification

- **Main difference is the target or response variable (y)**

# Regression vs Classification

- Main difference is the target or response variable (y)
- **One predicts real/floating-point values whereas the other predicts categorical values (predefined set)**

# Regression vs Classification

- Main difference is the target or response variable (y)
- One predicts real/floating-point values whereas the other predicts categorical values (predefined set)
- **Regression examples:**
  - **Diabetes scores**
  - **Healthcare cost**

# Regression vs Classification

- Main difference is the target or response variable (y)
- One predicts real/floating-point values whereas the other predicts categorical values (predefined set)
- Regression examples:
    - Diabetes scores
    - Healthcare cost

- **Classification examples:**
    - **Character recognition (10 classes)**
    - Yes/No (or Binary) questions:
        - Accept/reject loan application
        - Positive vs negative sentiment
    - Dog vs cat (2 classes), still binary class

# Regression vs Classification

- Main difference is the target or response variable (y)
- One predicts real/floating-point values whereas the other predicts categorical values (predefined set)
- Regression examples:
  - Diabetes scores
  - Healthcare cost

- **Classification examples:**
  - Character recognition (10 classes)
  - **Yes/No (or Binary) questions:**
    - **Accept/reject loan application**
    - **Positive vs negative sentiment**
  - Dog vs cat (2 classes), still binary class

# Regression vs Classification

- Main difference is the target or response variable (y)
- One predicts real/floating-point values whereas the other predicts categorical values (predefined set)
- Regression examples:
  - Diabetes scores
  - Healthcare cost

- **Classification examples:**
  - Character recognition (10 classes)
  - Yes/No (or Binary) questions:
    - Accept/reject loan application
    - Positive vs negative sentiment
  - **Dog vs cat (2 classes), still binary class**

# Regression vs Classification

- Main difference is the target or response variable (y)
- One predicts real/floating-point values whereas the other predicts categorical values (predefined set)
- Regression examples:
  - Diabetes scores
  - Healthcare cost
- We have learned about regression (not complete yet; will continue ..)

- Classification examples:
  - Character recognition (10 classes)
  - Yes/No (or Binary) questions:
    - Accept/reject loan application
    - Positive vs negative sentiment
  - Dog vs cat (2 classes), still binary class

- We will start our classification predictive modeling journey today
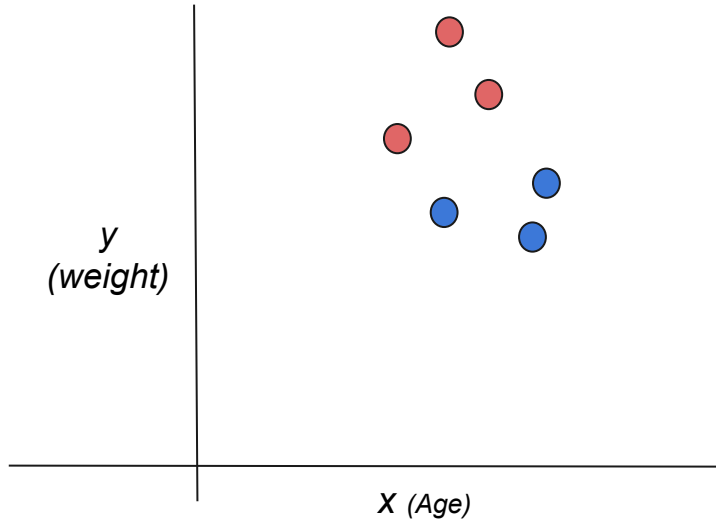
# Linear Line as a Threshold



*y (weight)*

*X (Age)*
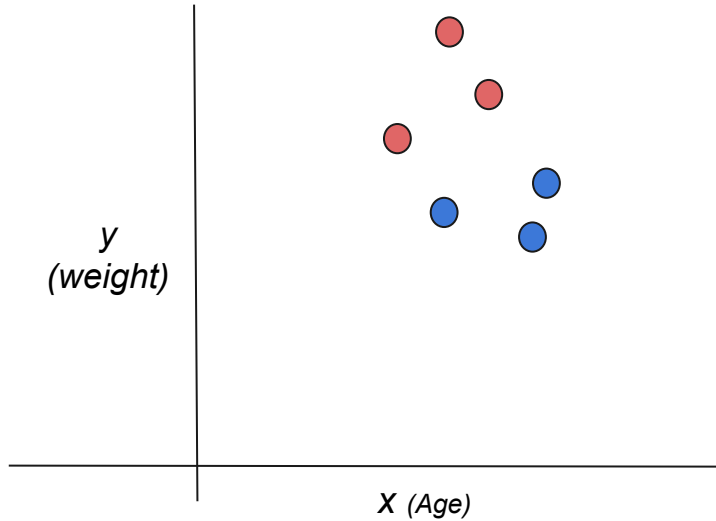
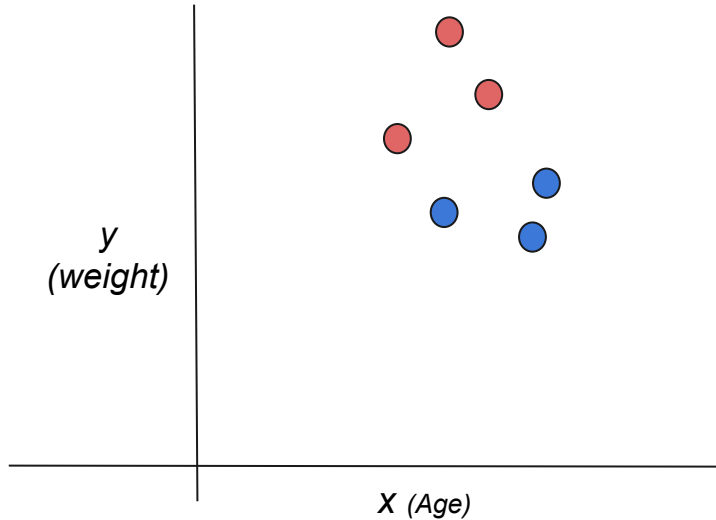- We have some overweight and underweight people

# Linear Line as a Threshold



*y*
*(weight)*

*X (Age)*
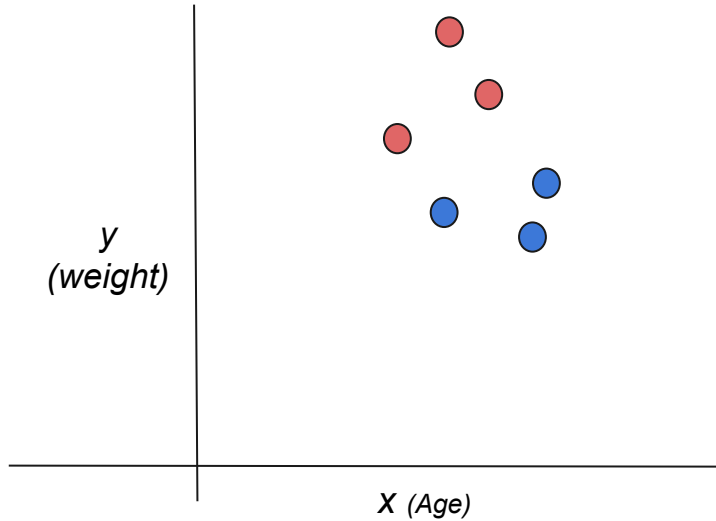
- We have some overweight and underweight people
- We need to learn a classifier

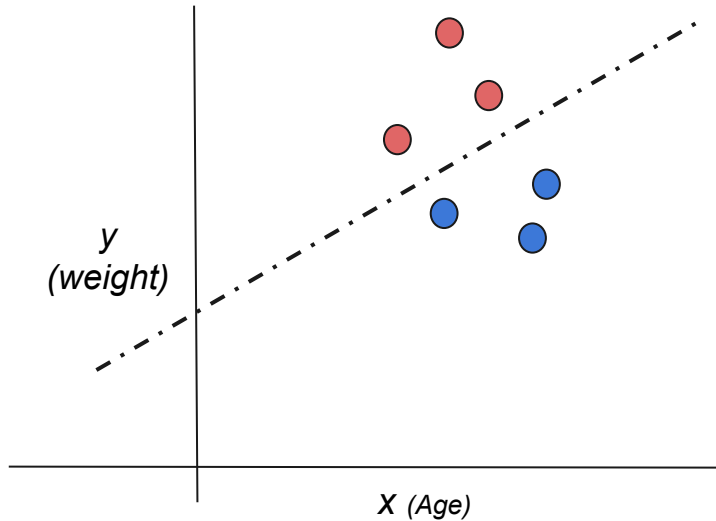# Linear Line as a Threshold



*y (weight)*

*X (Age)*

- We have some overweight and underweight people
- We need to learn a classifier
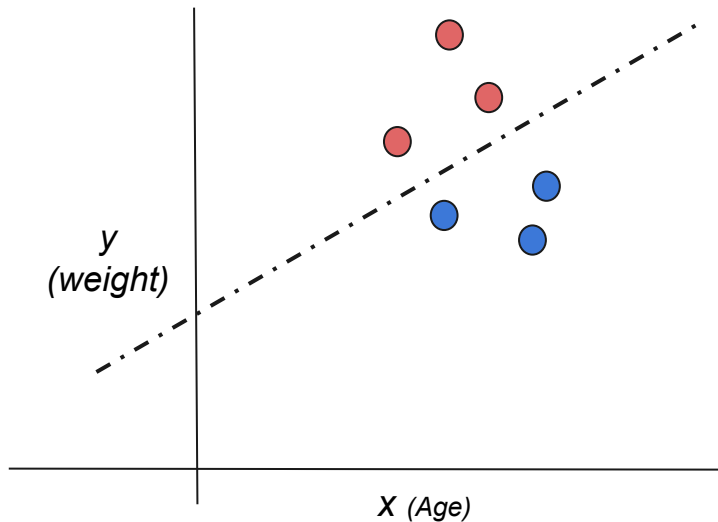- Can we do it using a straight line?

# Linear Line as a Threshold



*y (weight)*

*X (Age)*

- We have some overweight and underweight people
- We need to learn a classifier
- Can we do it using a straight line?
- Yes; how ..

# Linear Line as a Threshold



- We have some overweight and underweight people
- We need to learn a classifier
- Can we do it using a straight line?
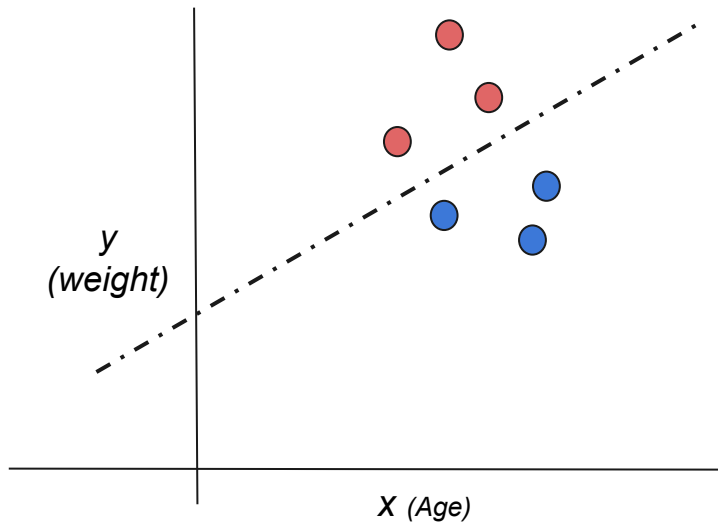- Yes; how ..

# Linear Line as a Threshold



y
(weight)

X (Age)

Straight Line as the Separator

$$y = \beta_0 + \beta_1 x$$

# Linear Line as a Threshold



y
(weight)

X (Age)

**Straight Line as the Separator**
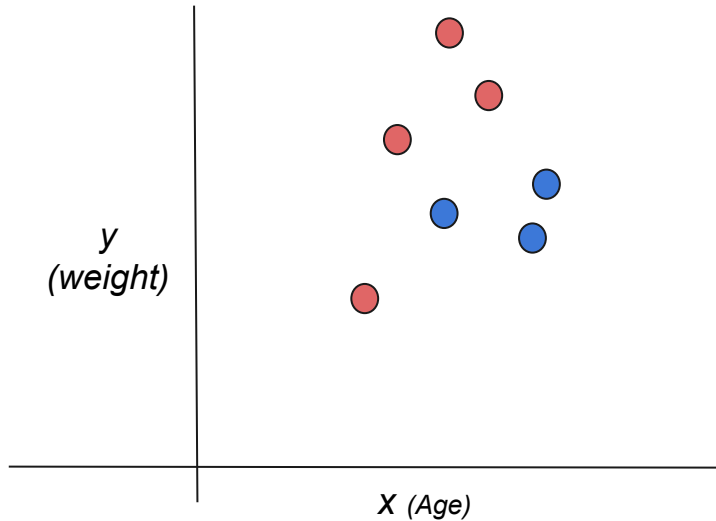
$$y = \beta_0 + \beta_1 x$$

**Classification Rule**

$$\hat{y} = \begin{cases} 1, & \text{if } \beta_0 + \beta_1 x > 0 \\ 0, & \text{otherwise} \end{cases}$$
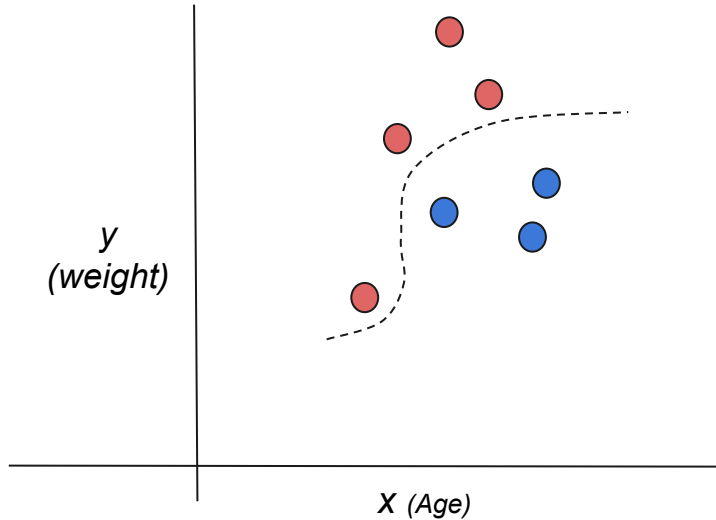
# Let's Try

# Linear Line as a Threshold



- We have some overweight and underweight people
- We need to learn a classifier
- Can we do it using a straight line?

# Linear Line as a Threshold



*y (weight)*

*X (Age)*

- We have some overweight and underweight people
- We need to learn a classifier
- Can we do it using a straight line?
- No; We will require a non-linear classifier.

# Classification Models

- Logistic Regression
- Random Forest Classifier
- Support Vector Machines (SVMs)
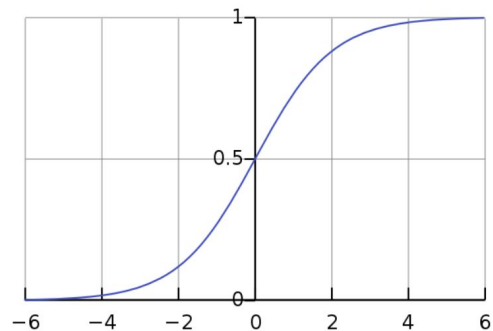- Boosting Classifiers
- Naive Bayes

# Logistic Regression
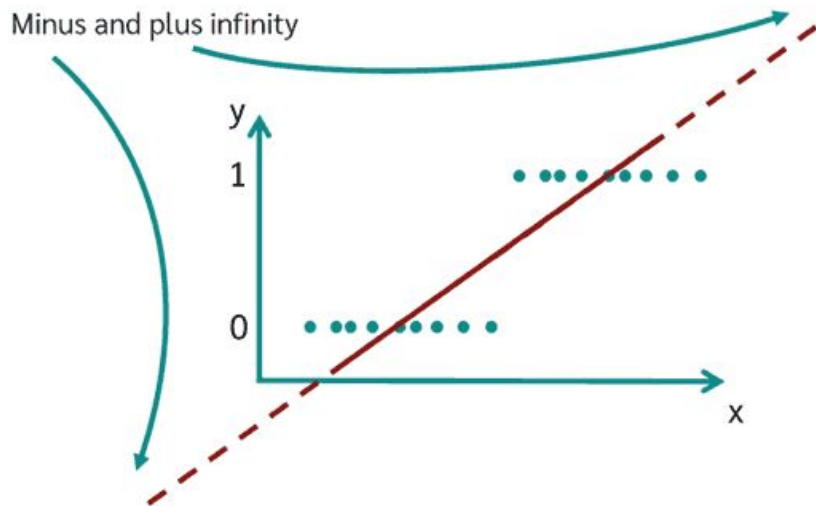
- Probabilistic classifier

  [Sigmoid function characteristic](#)
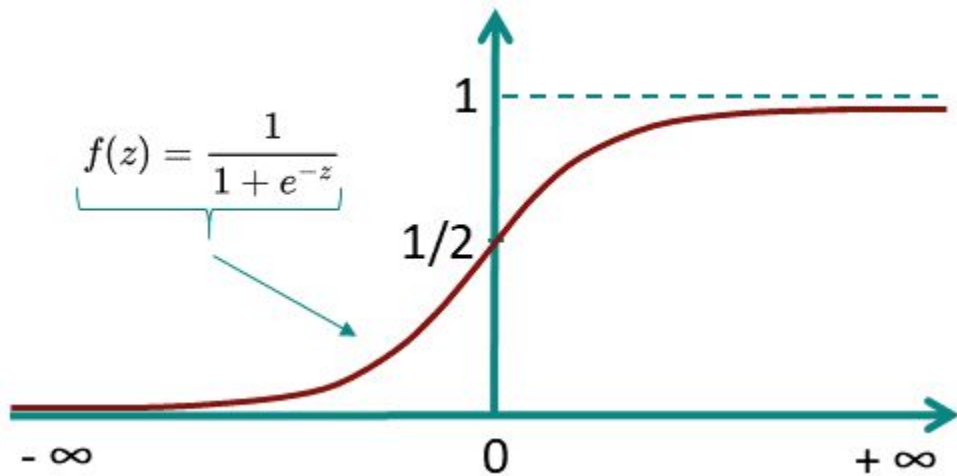
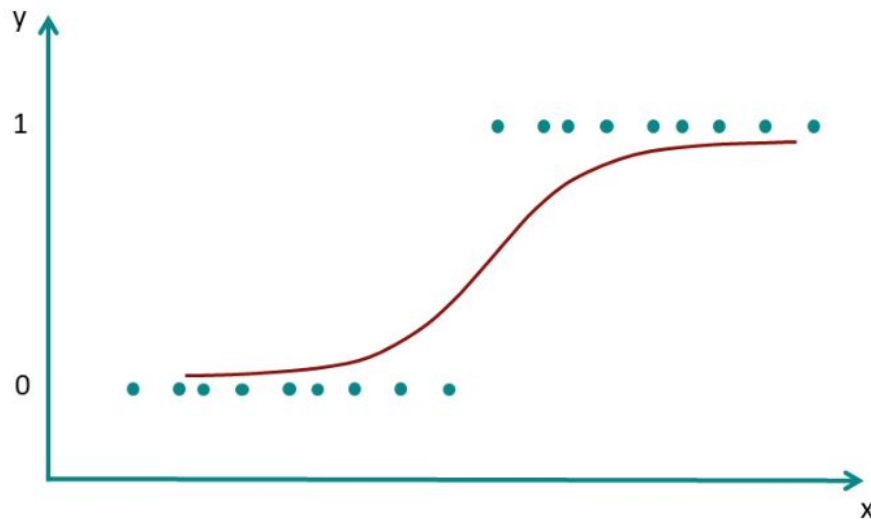$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

- Sigmoid function

# Logistic Regression



Minus and plus infinity

# Logistic Regression



$$f(z) = \frac{1}{1 + e^{-z}}$$

1

1/2

0

- ∞

+ ∞

src

# Logistic Regression

# QA