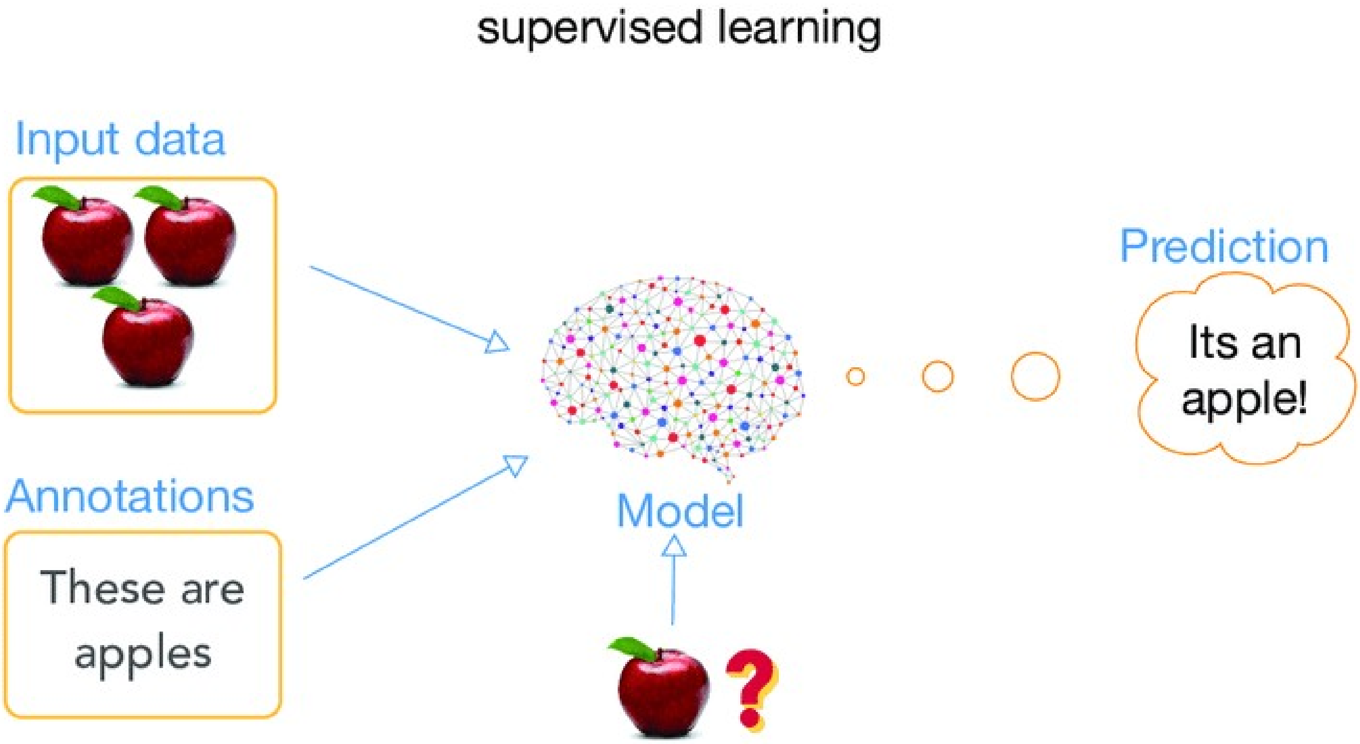
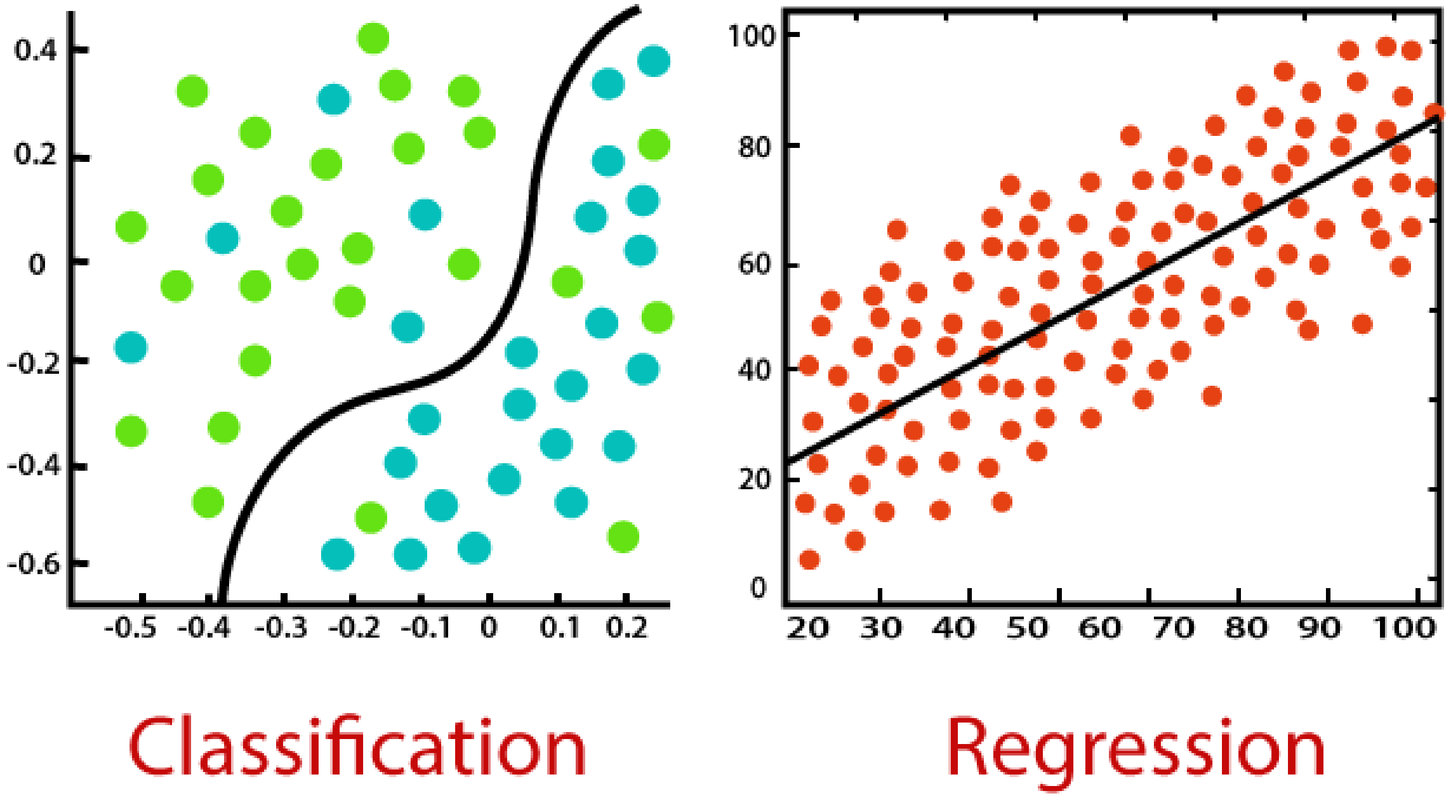
Supervised learning – logistic regression

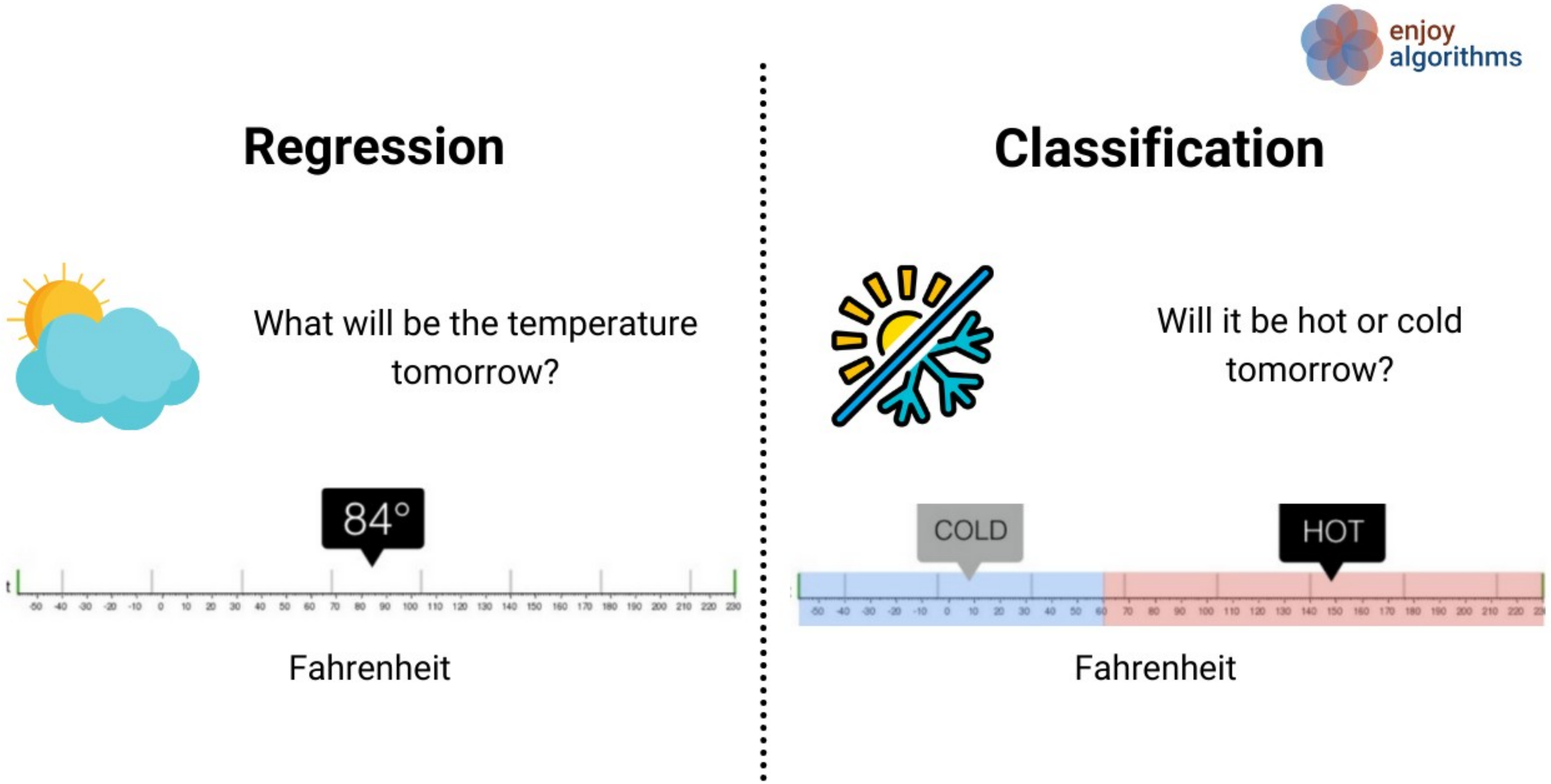
What is supervised learning?



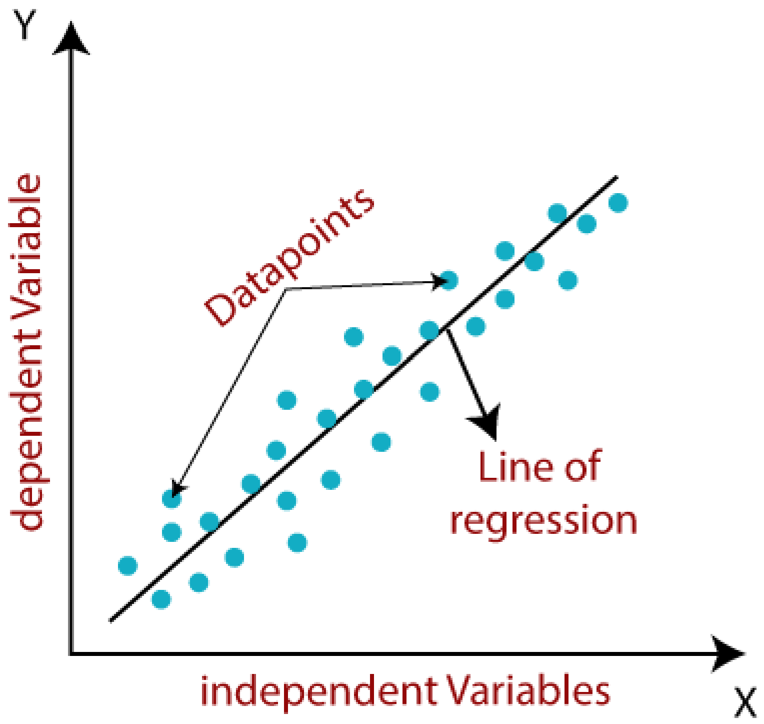
Regression vs Classification



Regression vs Classification

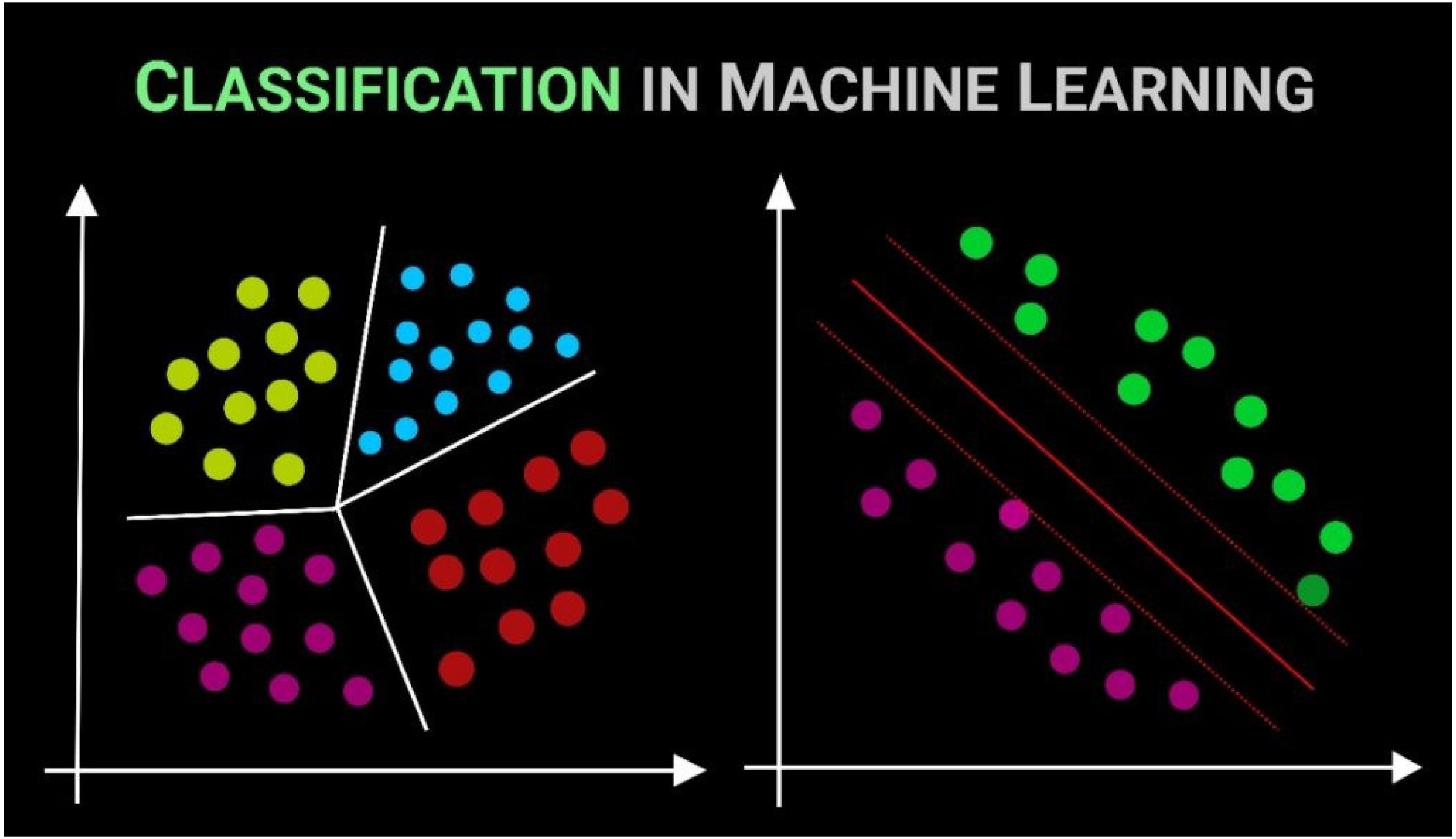


Regression – most well known case :)



$$Y = \beta_0 + \beta_1 X + \epsilon$$

Classification – assigning labels



Classification: What about qualitative outcomes?

- Consider an outcome measuring if a person voted in the latest election.

Classification: What about qualitative outcomes?

- Consider an outcome measuring if a person voted in the latest election.
  - We want to see if age is a factor in voting?

Classification: What about qualitative outcomes?

- Consider an outcome measuring if a person voted in the latest election.
  - We want to see if age is a factor in voting?
  - Can we use the simple linear regression framework?

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Y = 1 if person voted, 0 if not
X = age

Classification: Why not use linear regression?

- Let us extend the voting scenario a bit further.
  - Consider three outcomes for the persons vote – republican, democrat or independent.

Classification: Why not use linear regression?

- Let us extend the voting scenario a bit further.
  - Consider three outcomes for the persons vote – republican, democrat or independent.
    - How would you code this?

      Republican 1
      Democrats 2
      Independent 3

Classification: Why not use linear regression?

- Let us extend the voting scenario a bit further.
  - Consider three outcomes for the persons vote – republican, democrat or independent.
    - How would you code this?

      Republican 1
      Democrats 2
      Independent 3
    - What does this coding assume? What happens if we change the coding?

Classification: Why not use linear regression?

- Let us look at another aspect of the simple voting scenario (voted/not voted : Y=1/0).
  - Consider a linear model of this against age

$$Y = \beta_0 + \beta_1 X + \epsilon$$

  - Can Y be fit as anything other than 0 and 1?
    - How would you transform a fitted Y to 0 or 1?
    - What about less than 0 or more than 1?

Classification: So what can we use?

- Think of the dichotomous voting case (Y=1/0)
  - What should we be fitting?
    - The probability of an outcome
      - Disease
      - Voting
    - What function can we use?

Classification: Logistic regression

- One way to model the relationship between p(X) = P(Y=1 | X)

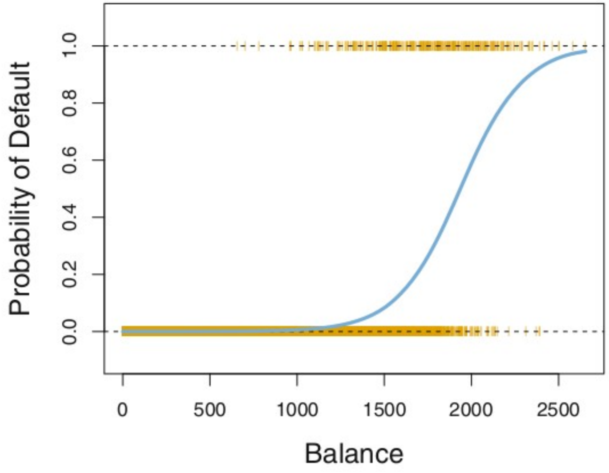
$$p(X) = P(Y = 1|X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}}$$

Classification: Logistic regression

- One way to model the relationship between p(X) = P(Y=1 | X)

$$p(X) = P(Y = 1|X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}}$$

- What is this monstrosity?

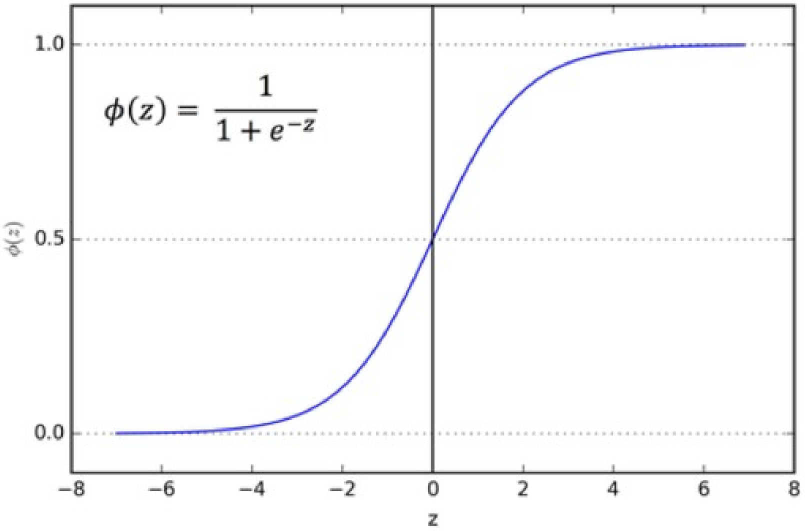


Classification: Logistic regression

- One way to model the relationship between p(X) = P(Y=1 | X)

$$p(X) = P(Y = 1|X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}}$$

- What is this monstrosity?
  - Sigmoid curve



Logistic regression

- Why is it still a linear model?

$$p(X) = P(Y=1|X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$$

$$\frac{p(X)}{1-p(X)} = e^{\beta_0+\beta_1 X}$$

$$log(\frac{p(X)}{1-p(X)}) = \beta_0 + \beta_1 X$$

Link function: Links Y
to the linear predictor
Log-odds or logit

Linear predictor:
Linear function of X

Logistic regression: What do the coefficients mean?

- Here is the model again:

$$log(\frac{p(X)}{1-p(X)}) = \beta_0 + \beta_1 X$$

- β_0 : Intercept – log-odds when X=0
- β_1 : Slope – Increase in log-odds with unit increase in X or $\exp(\beta_1)$ is the increase in odds for unit increase in X.

Logistic regression: Estimating the coefficients

- We can use a likelihood framework (cost function)
  - The likelihood formula is not analytically solvable
  - Use a numerical solver to estimate the coefficients
    - gradient descent
  - Use inbuilt methods in python

Logistic regression: More topics

- Multiple logistic regression
- Polytomous regression

Exercise Time