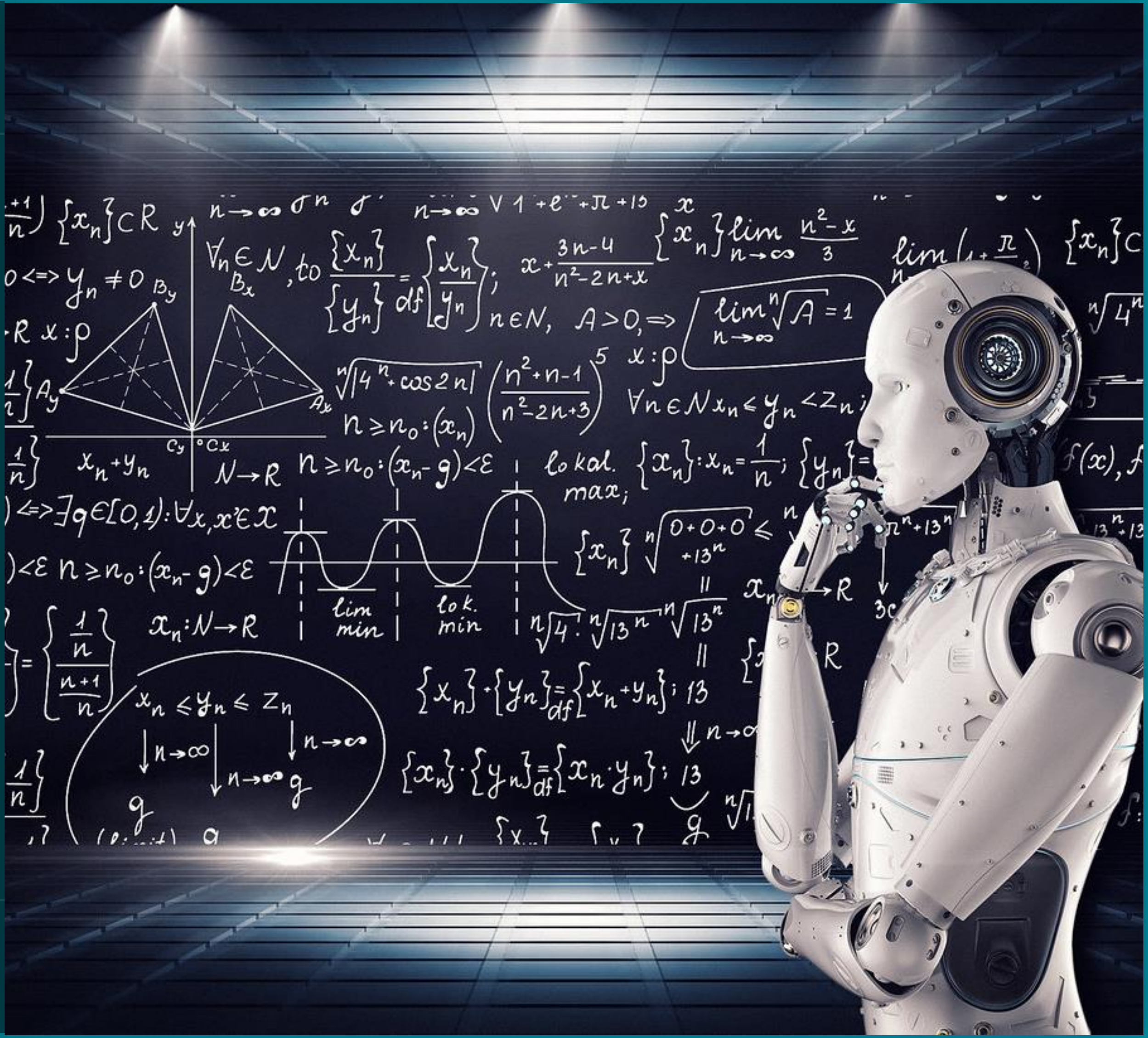
Introduction to Machine Learning

Supervised learning



HOSPITAL DA LUZ
LEARNING HEALTH
TRAINING, RESEARCH & INNOVATION CENTER

Summary

26 Jun

- Introduction to supervised learning
- Linear regression
  - Simple linear regression
  - Multiple linear regression
  - Gradient descent

27 Jun

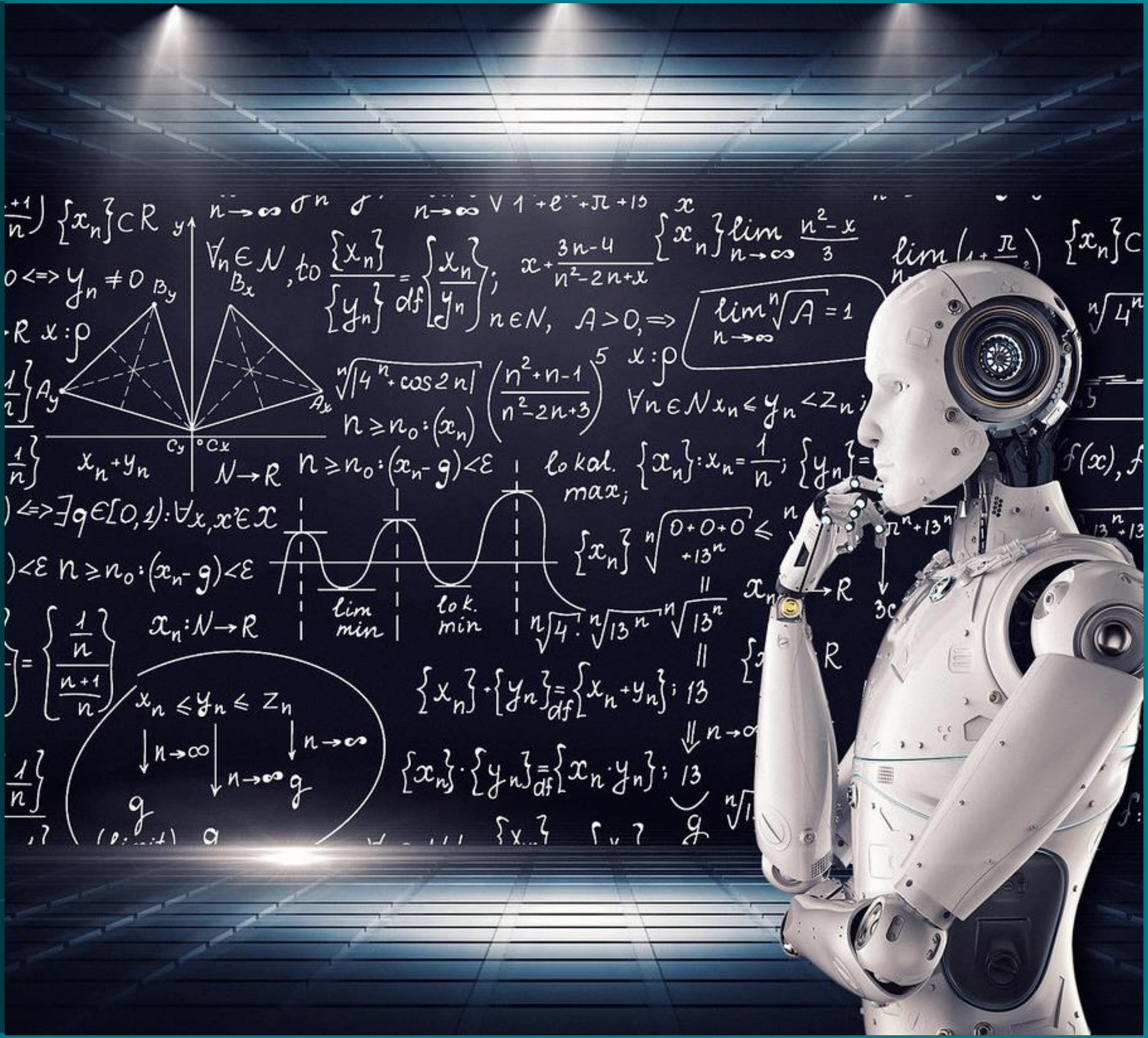
- Classification
  - Logistic regression
  - KNN

2 Jul

- Non-linear world
  - Polynomials and variants
  - Tree-based methods

Introduction to Machine Learning

Introduction to Supervised Learning



HOSPITAL DA LUZ
LEARNING HEALTH
TRAINING, RESEARCH & INNOVATION CENTER

Machine Learning Types

Supervised Learning

Unsupervised Learning

Reinforcing Learning

Supervised Learning

Input $\qquad\qquad$ Output

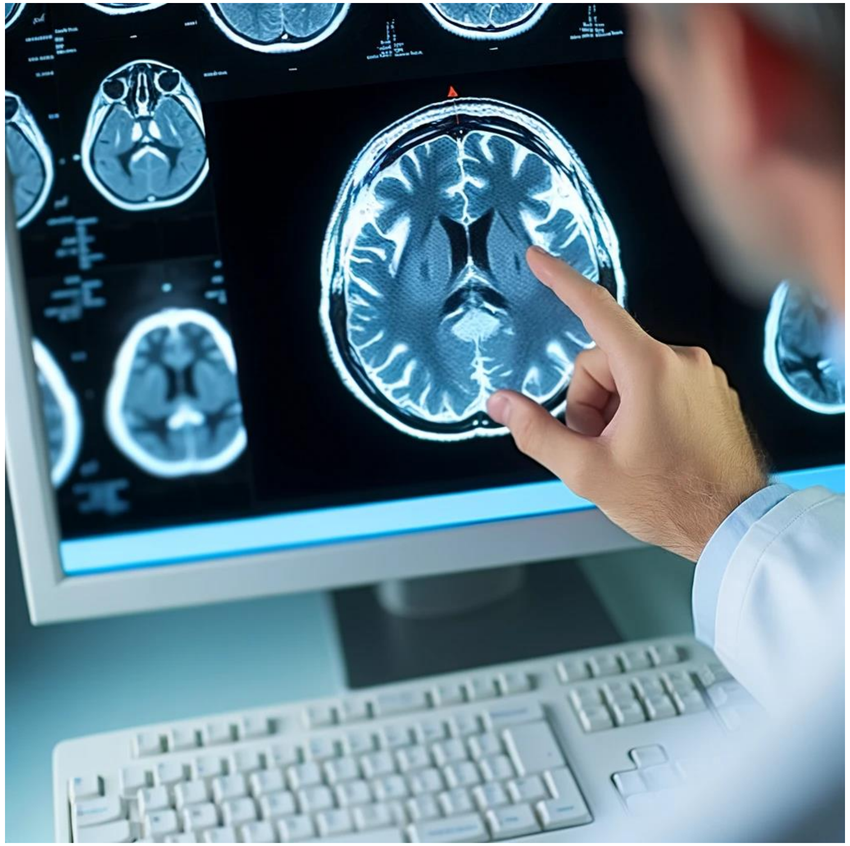
$$X \xrightarrow[f(X)]{\qquad\qquad\qquad} Y$$

Clinical applications of supervised learning

Input *X*: CT scans

Output *Y:* Tumor presence

Application *f(X)*: Image analysis

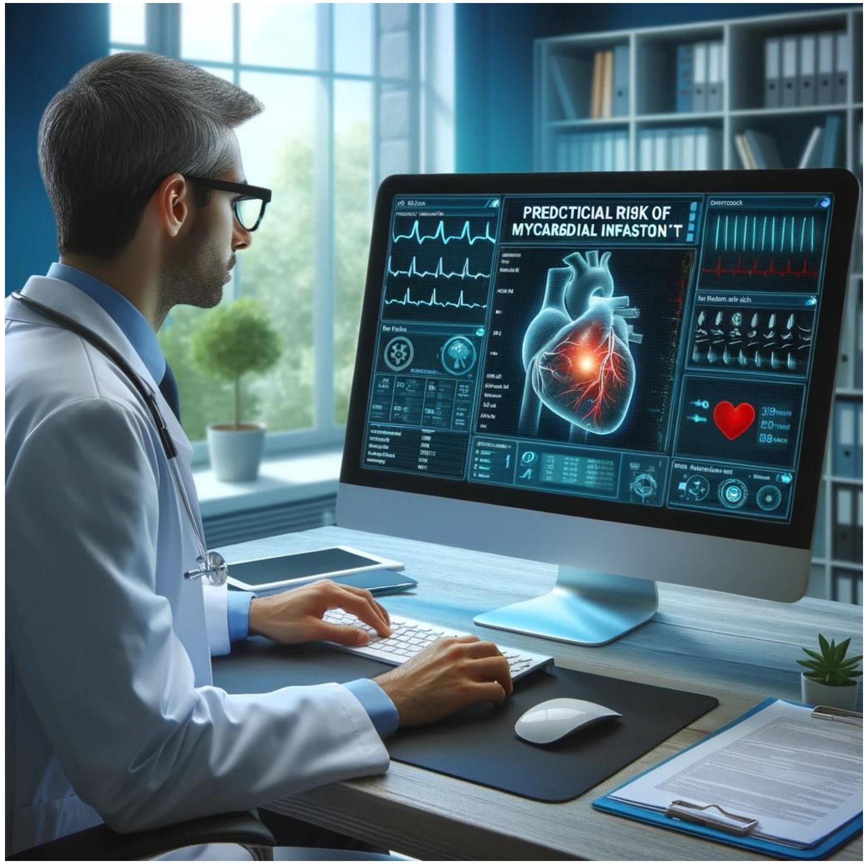


Clinical applications of supervised learning

Input X: Clinical data

Output Y: Myocardial infarction

Application f(X): Clinical risk prediction

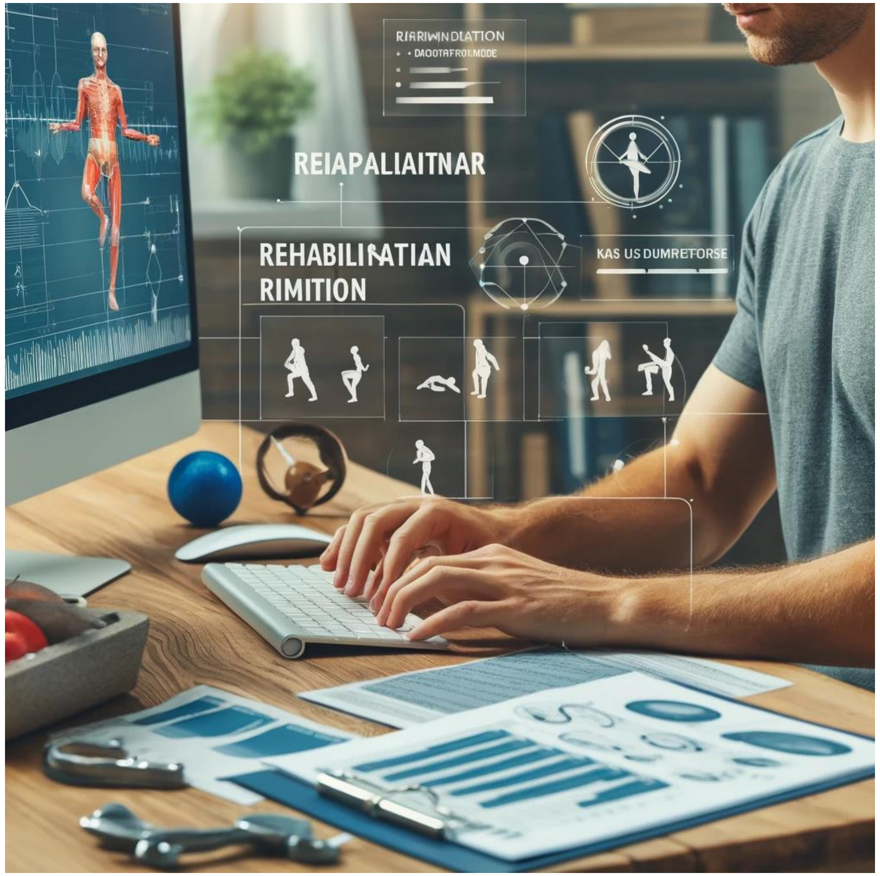


Clinical applications of supervised learning

Input X: Physical therapy assessment

Output Y: Improvement in physical function

Application f(X): Planning rehabilitation regimens



Clinical applications of supervised learning

Input X: Chemotherapy, patient data

Output Y: Tumor growth

Application f(X): Predict tumor response

Why estimate $f(X)$?

- Prediction

$$\hat{Y} = \hat{f}(X) + \varepsilon$$

- Inference

$$Y = f(X) + \varepsilon$$

How to estimate *f*(*X*)?

(x_1,y_1)

(x_3,y_3)

(x_2,y_2)

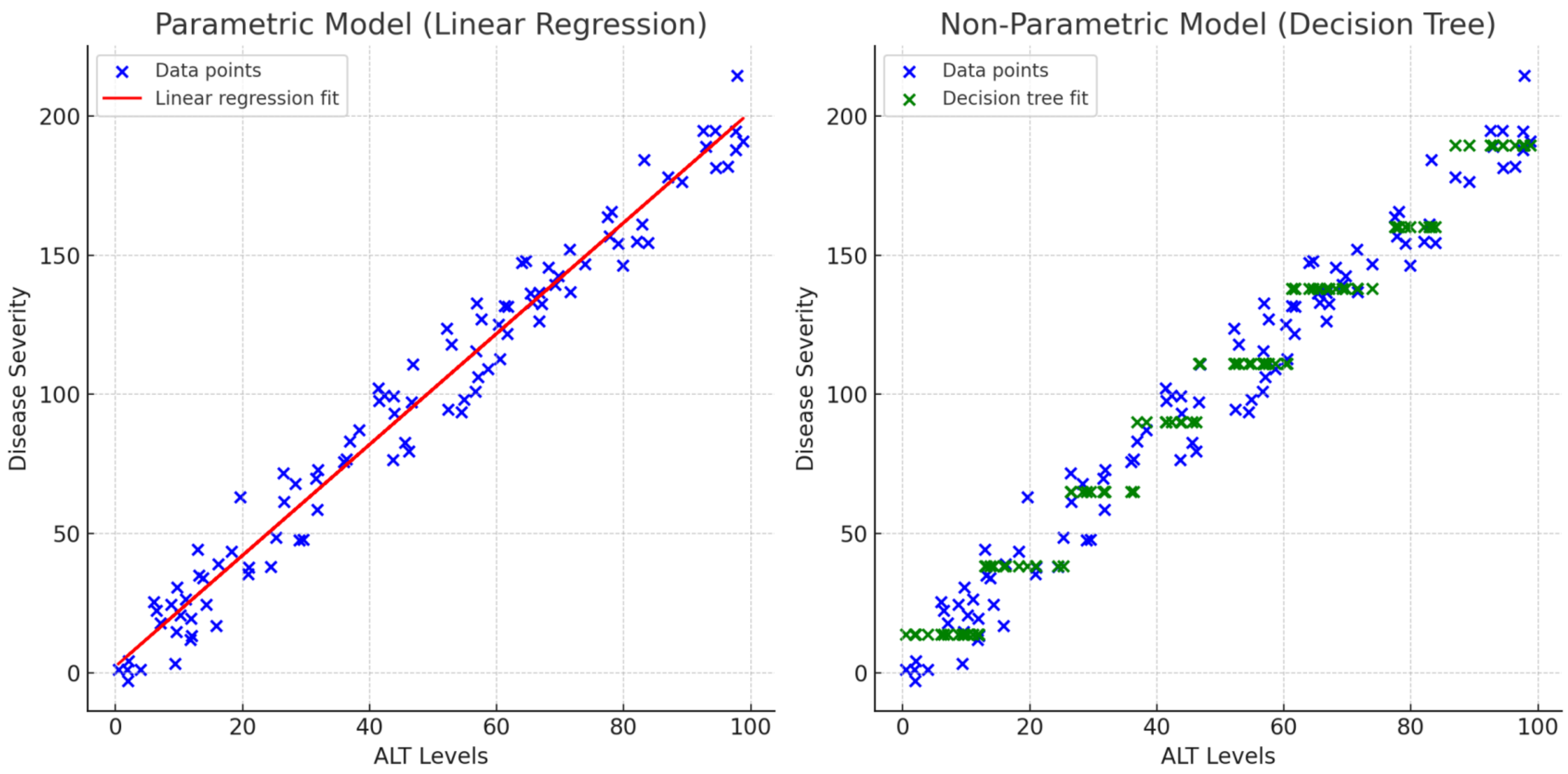
(x_n,y_n)

$$Y \approx f(X)$$

How to estimate *f*(*X*)?

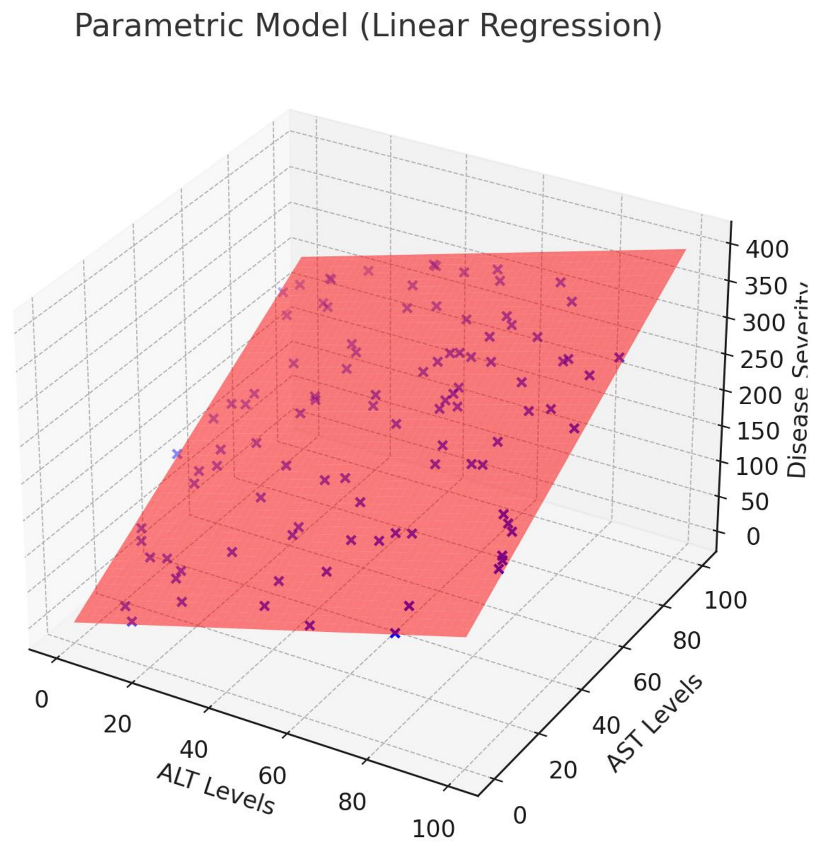
- *X*
  - Independent variable(s), inputs, predictors, features
- *Y*
  - Dependent variable, output, response, outcome
  - **Regression** if *Y* is continuous
  - **Classification** if *Y* is binary
- Parameters of *f(X)*

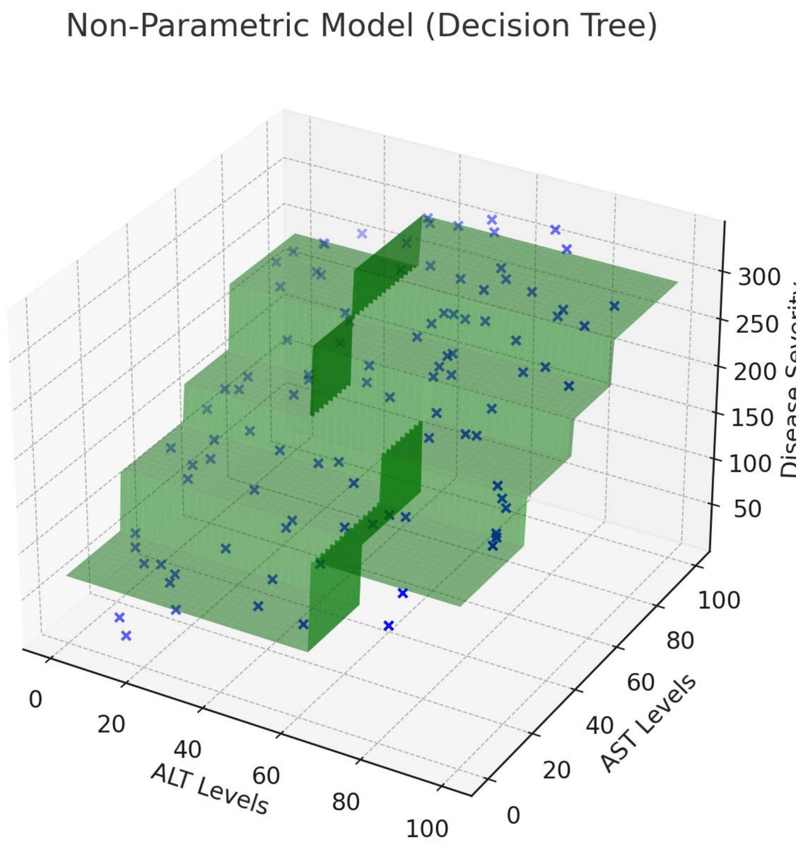
How to estimate *f*(*X*)?



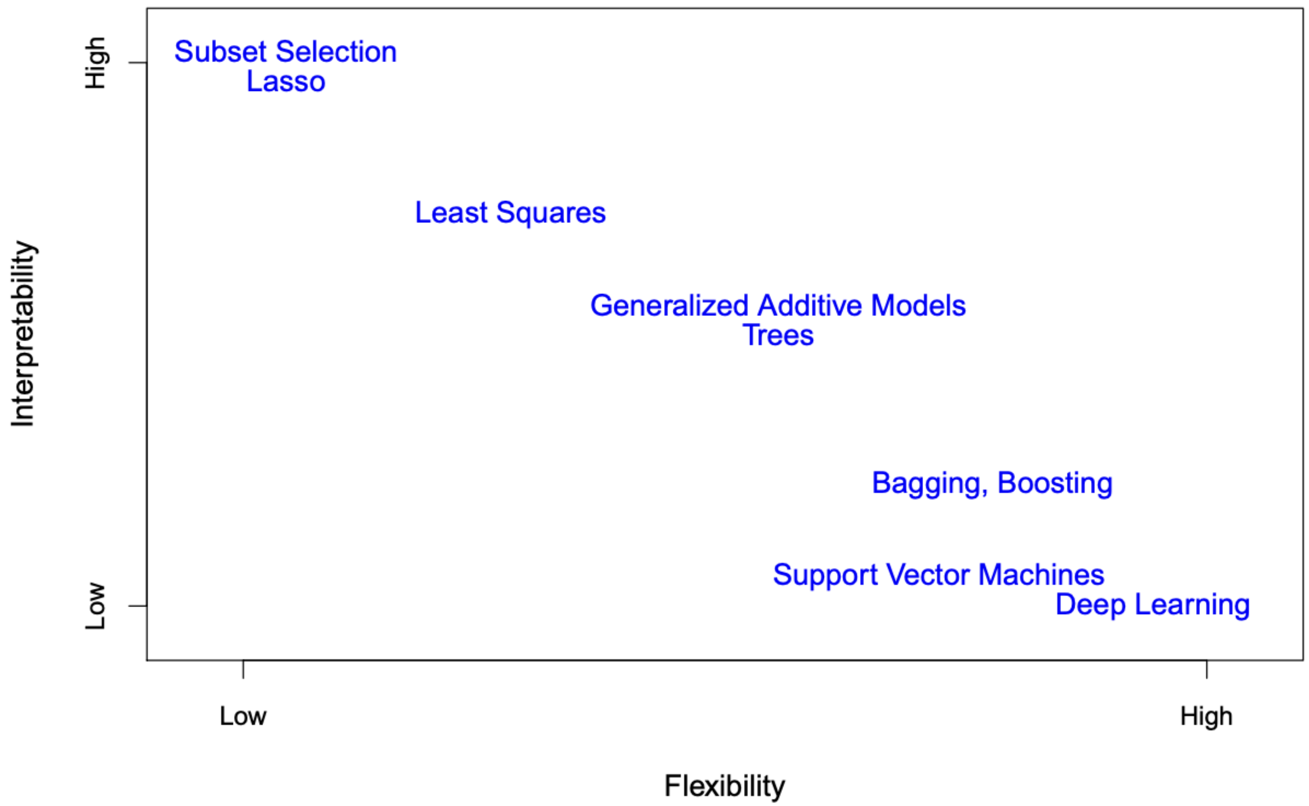
HOSPITAL DA LUZ LEARNING HEALTH
TRAINING, RESEARCH & INNOVATION CENTER

How to estimate *f*(*X*)?





How to estimate *f*(*X*)?



How to evaluate estimation of *f*(*X*)?

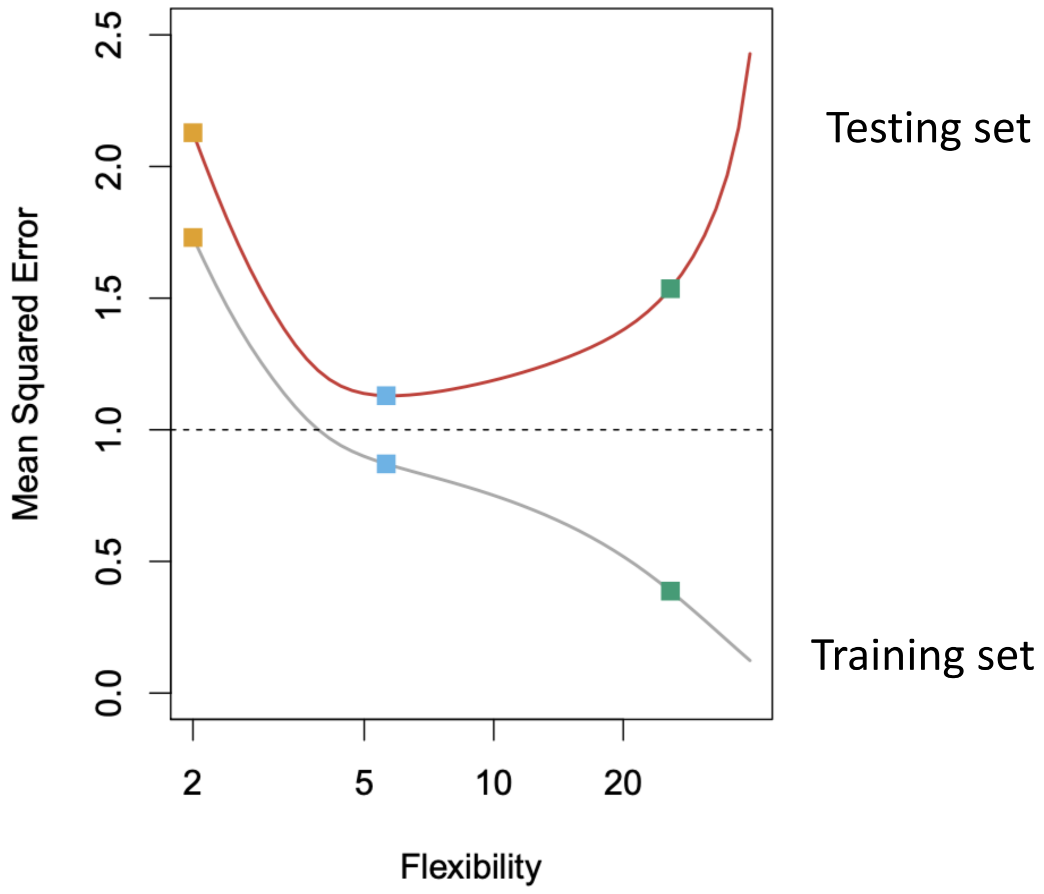
We assess the error

$$Y \approx f(X)$$

$$J(y, \hat{y}) = MSE = \frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2$$

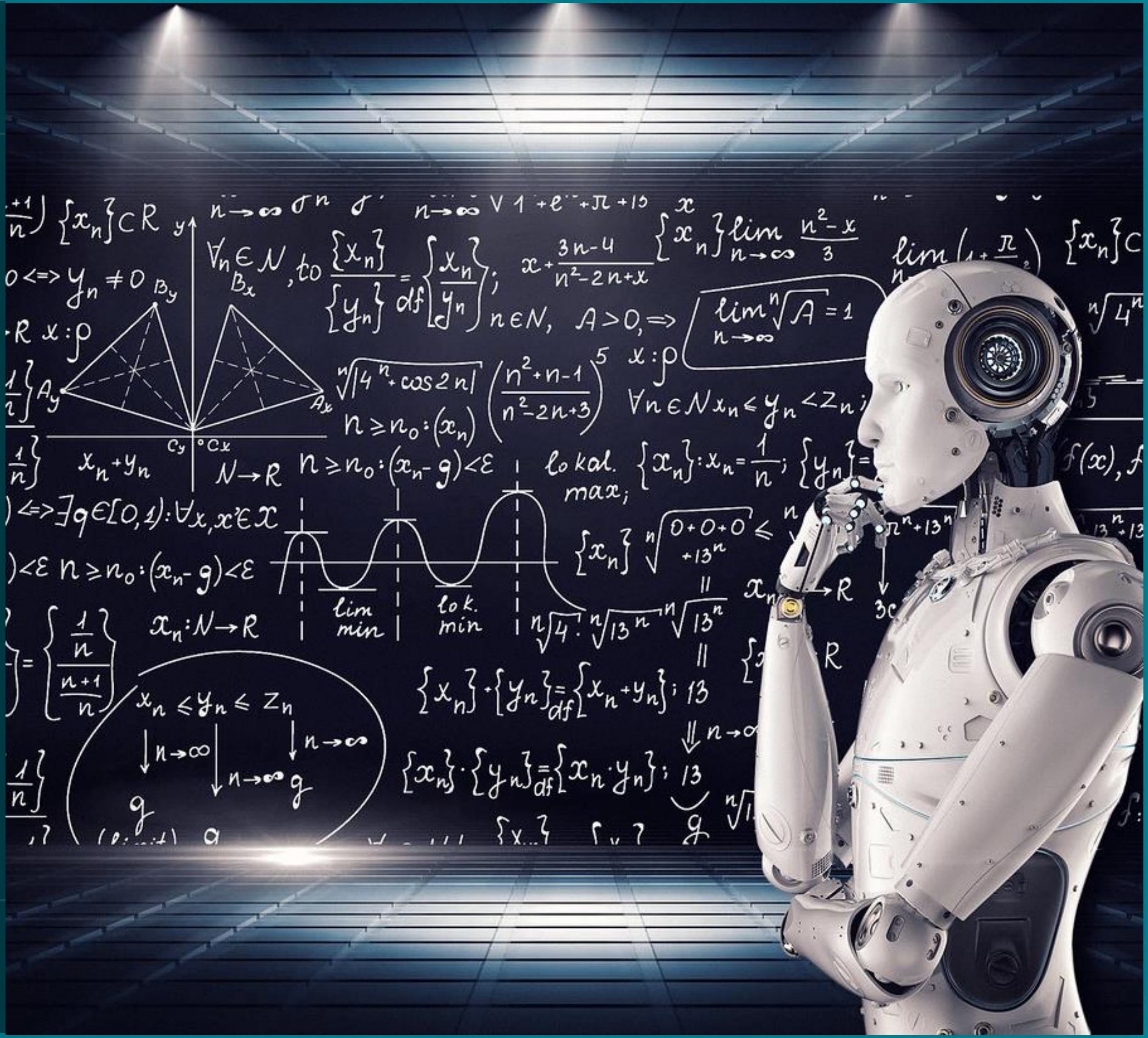
Cost function

How to evaluate estimation of *f*(*X*)?



Introduction to Machine Learning

Linear regression



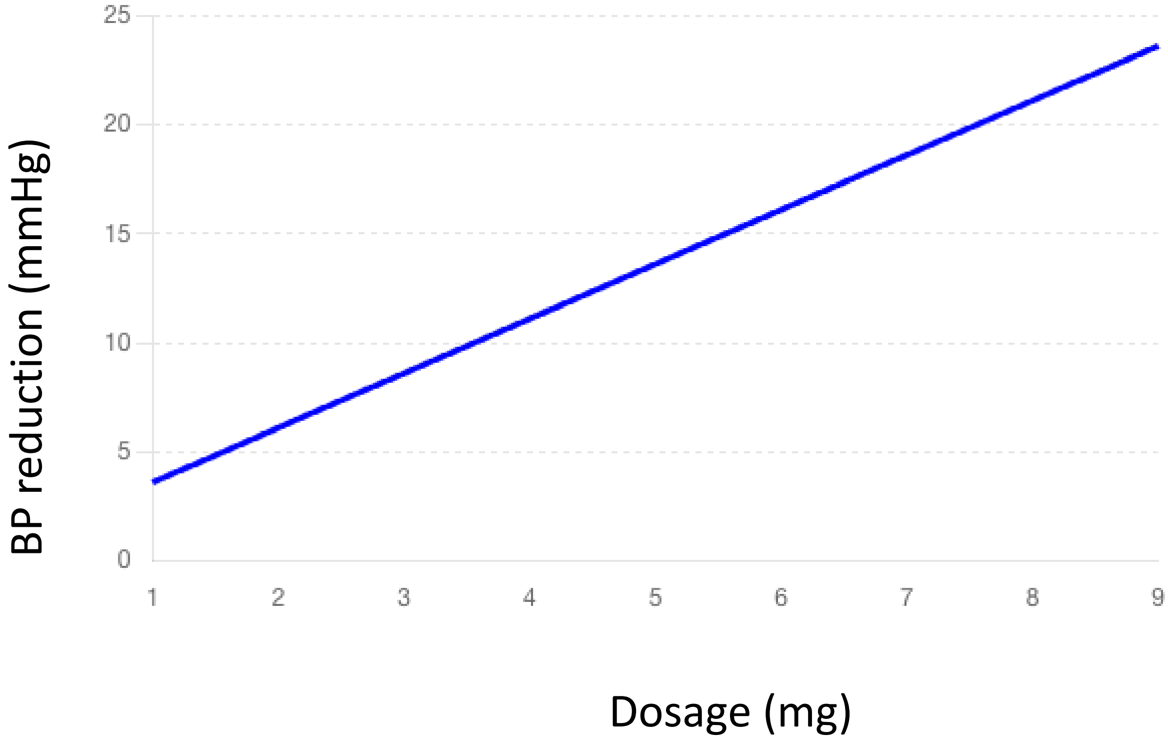
Simple Linear regression

$$\hat{Y} \approx \beta_0 + \beta_1 X$$

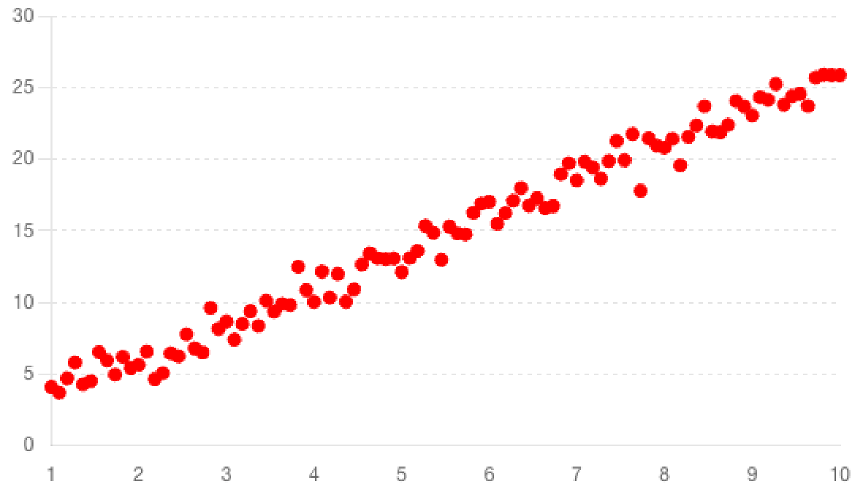
Linear regression

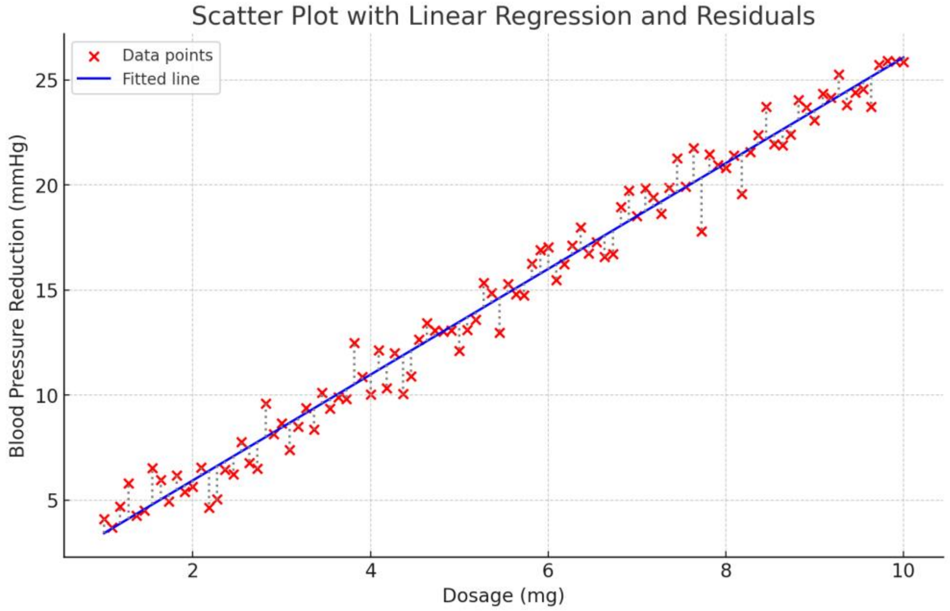
$$\hat{Y} \approx \beta_0 + \beta_1 X$$

BloodPressureReduction = 4.1 + 2.5 * Dosage



Fitting data





$$e_i = y_i - \hat{y}_i$$

$$\mathrm{SSR} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 X_i))^2$$

HOSPITAL DA LUZ LEARNING HEALTH
TRAINING, RESEARCH & INNOVATION CENTER

Evaluating a Linear regression: assessing model's fit

- Mean Square Error (MSE)
  - MSE is the average of the squared differences between the actual and the predicted values

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Lower MSE indicates a better fit

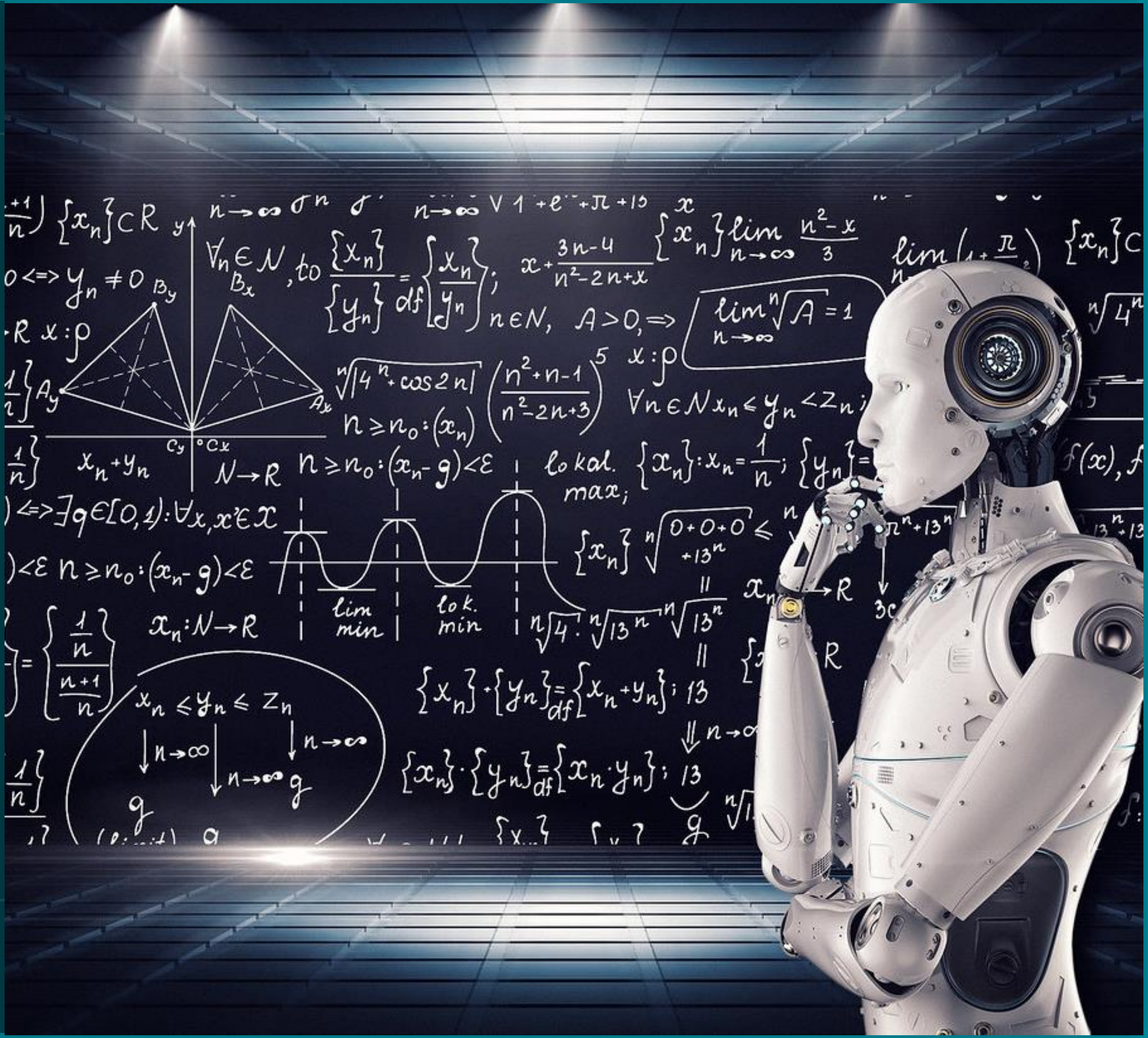
- R-squared (R^2)
  - Measures the proportion of the variance in the dependent variable that is predictable from the independent variable(s)

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Higher R^2 indicates a better fit

Introduction to Machine Learning

Multiple Linear regression

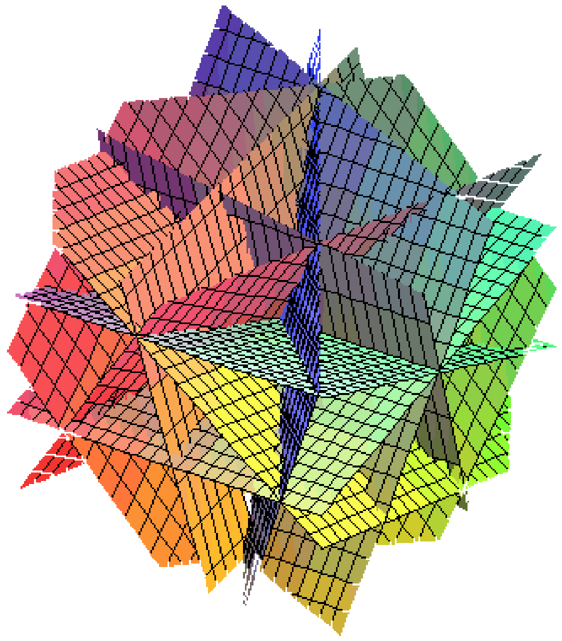


HOSPITAL DA LUZ
LEARNING HEALTH
TRAINING, RESEARCH & INNOVATION CENTER

Multiple linear regression

$$\hat{y} = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \beta_4 \cdot x_4 + \beta_5 \cdot x_5$$

$$\hat{y} = \beta_0 + \sum_{i=1}^{N} \beta_i \cdot x_i$$



HOSPITAL DA LUZ LEARNING HEALTH
TRAINING, RESEARCH & INNOVATION CENTER

Evaluating a multiple linear regression

- Mean Square Error (MSE)
  - MSE is the average of the squared differences between the actual and the predicted values

$$J(y, \hat{y}) = MSE = \frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2$$

Lower MSE indicates a better fit

- R-squared (R^2) or coefficient of determination
  - Measures the proportion of the variance in the dependent variable that is predictable from the independent variable(s)

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Higher R^2 indicates a better fit

Linear regression: recap

- **Pros**
  - Really easy to understand (comparing to other algorithms)
  - Fast optimization (comparing to other algorithms)
  - Easy to extend the model (see next lessons)

- **Cons**
  - Sensible to outliers
  - Assumes that there is no multicollinearity
  - Feature scaling is required
  - Monotonicity assumption: for the model, the relation between each feature and the output