Statistical learning

Ngô Minh Nhựt

2025

Course info

- Materials
  - Lecture notes
  - Books:
    - Kevin P. Murphy, Probabilistic Machine Learning: An Introduction, MIT Press, March 2022
    - G. James, An Introduction to Statistical Learning (2nd edition), Springer, 2021
    - C. Bishop, Pattern Recognition and Machine Learning, Springer, 2006 (pdf)
- Lecturers
  - Theory: Ngô Minh Nhựt
  - Labs: Lê Long Quốc

Course info

- Labs and coding environment
  - Through instruction session and Q&A on forum
  - Python (with Visual Code)
  - Google Colab (for training with powerful machines)

https://code.visualstudio.com

https://colab.research.google.com

Course info

- Tests
  - Quiz and report: 10%
  - Lab assignment: 15%
  - Mid-term test: 35%
  - Final project: 40%

Course info

- This course provides
  - basic knowledge about machine learning and
  - underlying mechanism about machine learning as well as
  - some applications in practical problems

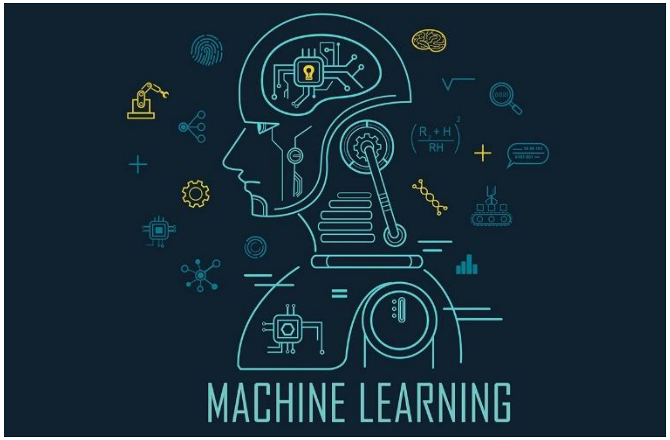
Course topic

Week	Topic	Week	Topic
1	Introduction	8	Support vector machine
2	Linear regression	9	K-mean
3	Logistic regression	10	Dimensionality reduction
4	Overfitting and model validation	11	Decision tree
5	Neural network		
6	Midterm test		
7	Convolutional neural network		

Machine learning

❑ A machine learning algorithm is an algorithm learning to accomplish a task by observing data.

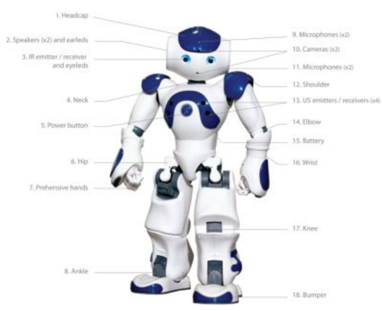
- Used on complex tasks where it's hard to develop algorithms with handcrafted-rules
- Exploits patterns in observed data and extract rules automatically

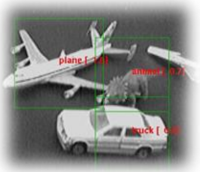


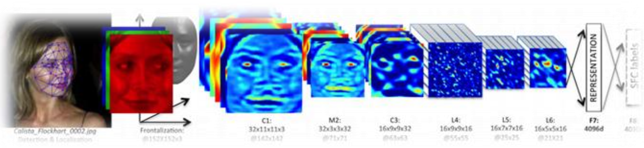
Applications

- Computer vision
- Speech to text, text to speech
- Financial analysis
- Self-driving cars
- Ads-targeting
- Virtual assistant

Hi, I am listening…







Example: object detection

- Learn from annotated corpus of examples (a dataset) to classify unknown images among different object types
- Observe images to learn patterns
- Lot of data available (i.e: ImageNet dataset)
- Very good error rates (< 5% with deep-CNN)

Types of machine learning algorithms

❑ Supervised

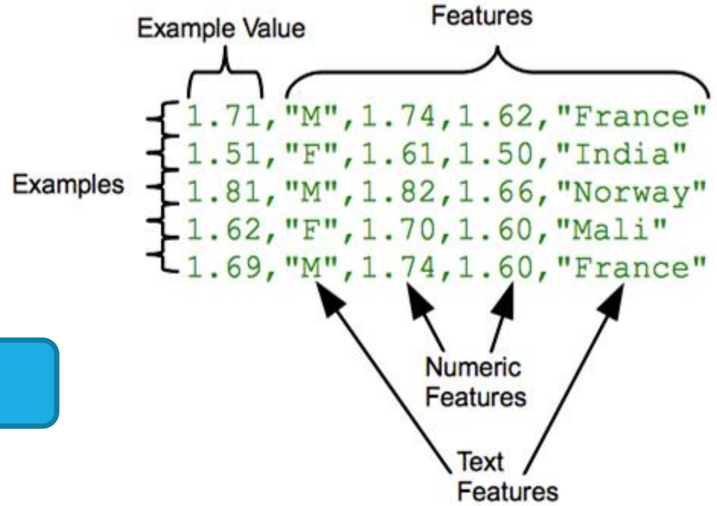
- Learn a function by observing labeled samples, containing input and expected output
- Classification
- Regression

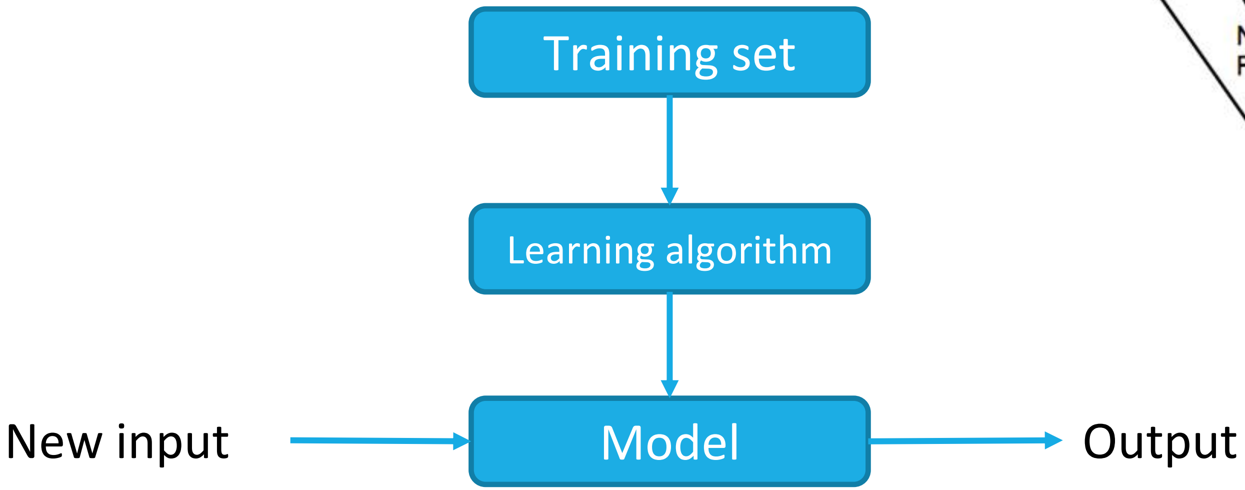
❑ Unsupervised

- Find underlying relations in data by observing unlabeled samples, containing only input
- Clustering
- Dimensionality reduction

Machine learning

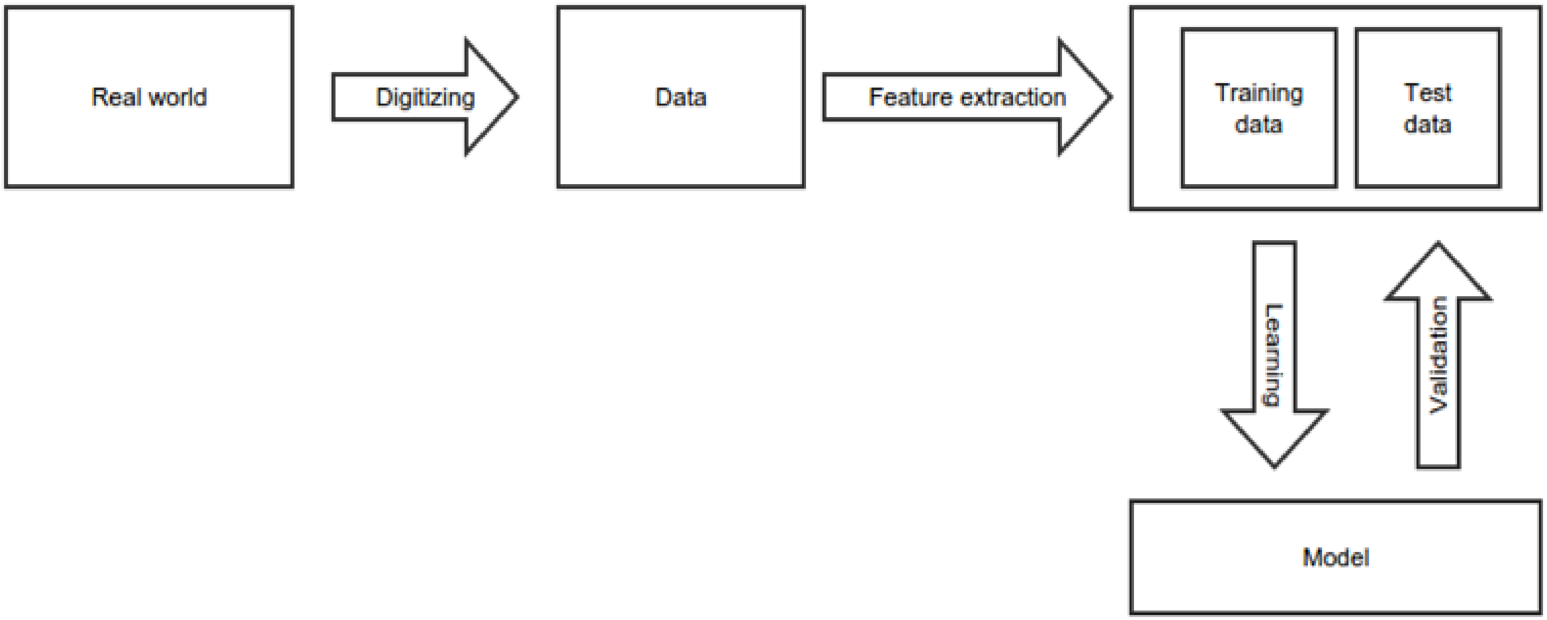
- Training set





Machine learning

- Learning process



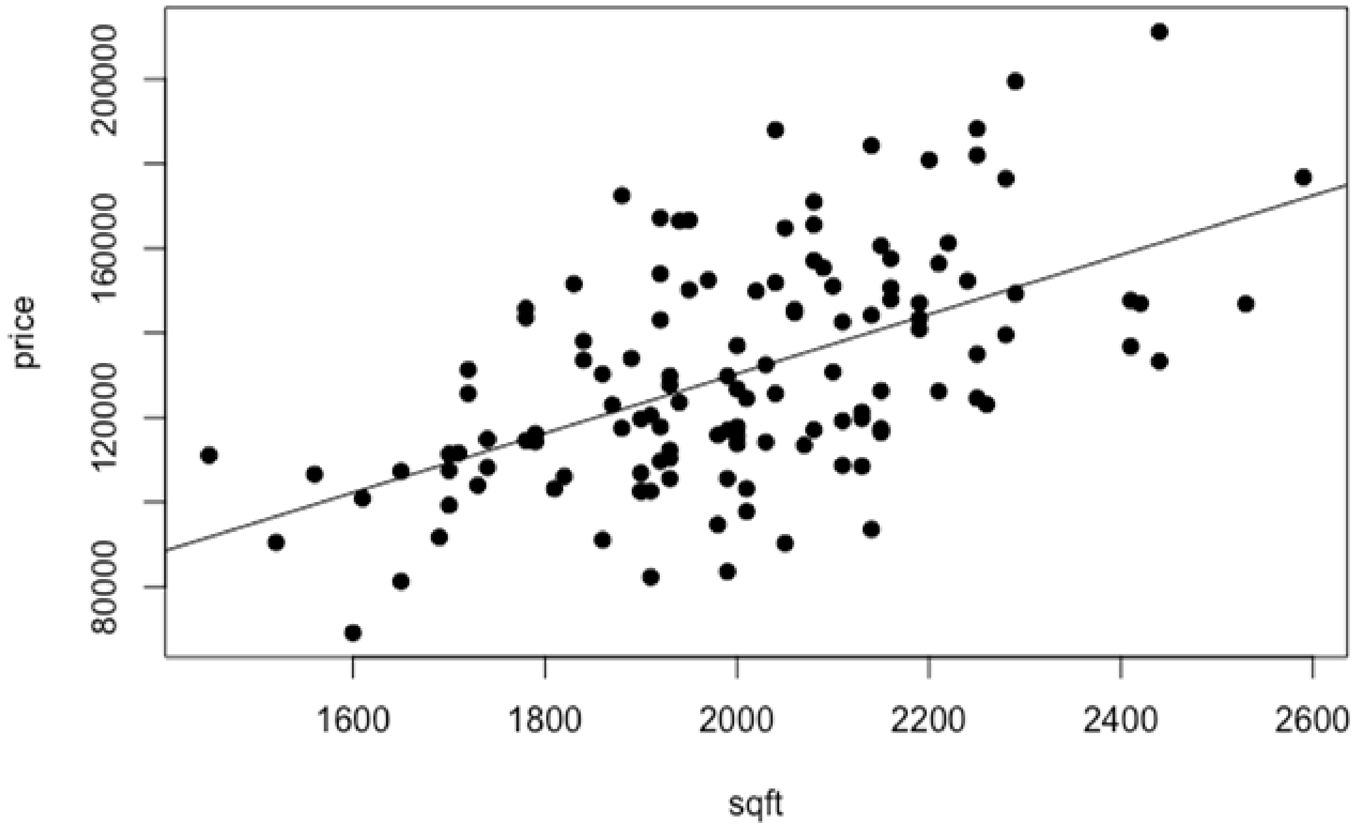
Feature extraction

- Extract the information useful for the task (features) we want to learn
- Example
  - Stock market time-series → [opening price, closing price, lowest price, highest price]
  - Image → Image with edges filtered
  - Document → bag-of-word

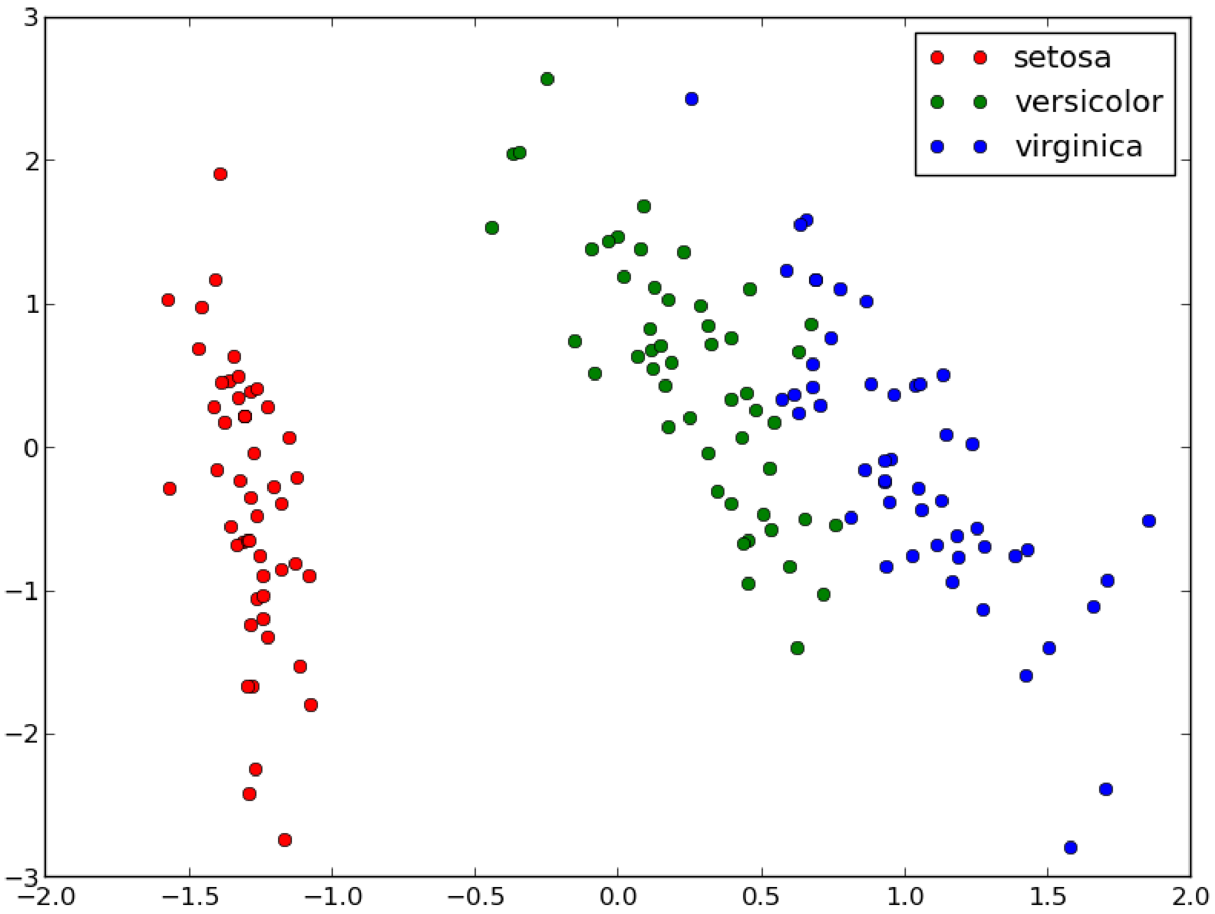
Classification vs Regression

- Regression: learn a function mapping an input to a real value
  - E.g., predict the temperature of tomorrow given some meteo signals
- Classification: learn a function mapping an input element to a class (within a finite set of possible classes)
  - E.g., predict tomorrow weather: {sunny, cloudy, rainy} given some meteo signals

Regression

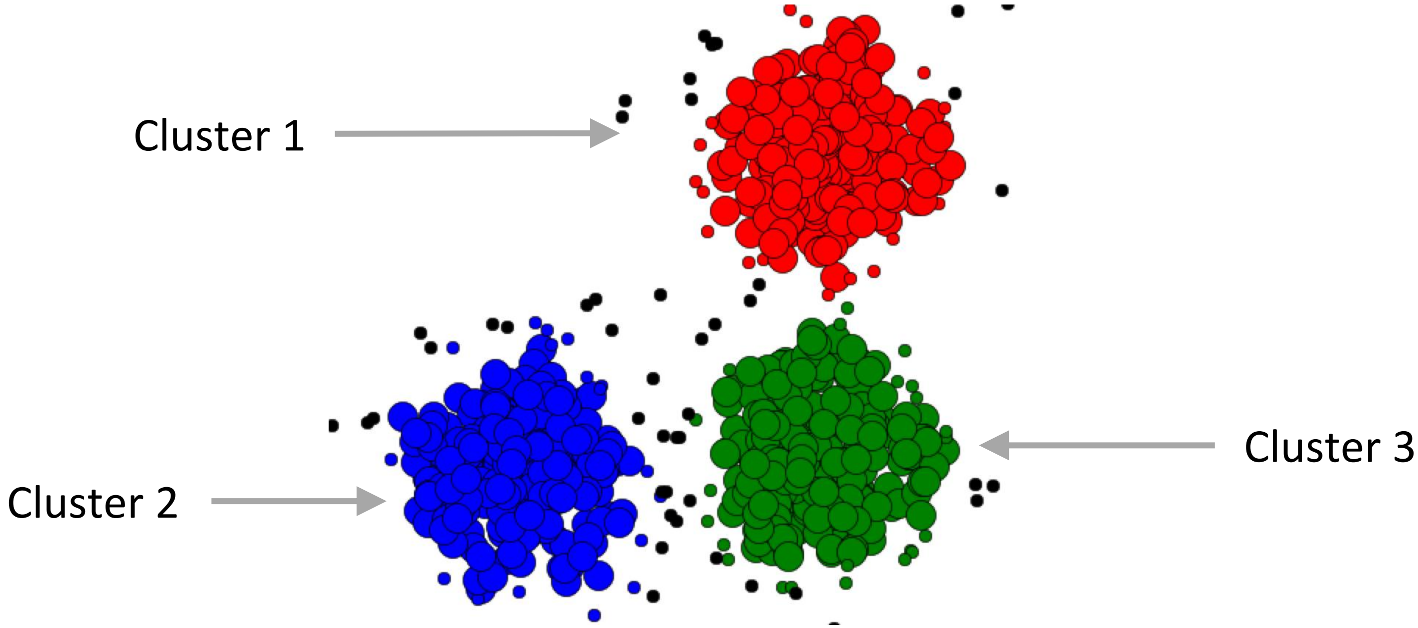

price
80000
120000
160000
200000
1600
1800
2000
2200
2400
2600
sqft

Classification



Clustering

- Separate different observed data points in similar groups (clusters). Observed data are unlabeled.



Linear algebra review

- Matrix and vector
- Summation and dot product
- Matrix multiplication
- Inversion and transposition

Matrix and vector

- Matrix is a two-dimension array, e.g.,

$$A = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$$

  - A is a 2x3 matrix consisting of 2 rows and 3 columns

- Vector is a matrix with multiple rows and one column

$$B = \begin{bmatrix} x \\ y \\ w \\ z \end{bmatrix}$$

  - B is a vector or 4x1 matrix

Matrix and vector

- Unit matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

  - Square matrix
  - Diagonal from left-top to right-bottom are all one
  - Other elements are zero
  - Notation of unit matrix: I

Matrix and vector

- Notations and symbols
  - A_{ij}: element at i-th row and j-th column of matrix A
  - Vector with n dimensions is a vector with n rows
  - v_i: element at i-th row of vector v
  - Matrix is represented by upper case characters, vector by lower case ones
  - Scalar is a quantity having one value
  - $\mathbb{R}$: set of real scalars
  - $\mathbb{R}^n$: set of n-dimension vectors

Summation and dot product

- Summation of matrixes is done element wise, e.g.,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} x & y \\ w & z \end{bmatrix} = \begin{bmatrix} a+x & b+y \\ c+w & d+z \end{bmatrix}$$

  - Dimension of output matrix does not change

- Multiplication of matrix and a scalar

$$\alpha \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \alpha a & \alpha b \\ \alpha c & \alpha d \end{bmatrix}$$

Matrix multiplication

- Multiplication of matrix and vector

$$\begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} * \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \\ ex + fy \end{bmatrix}$$

  - Output is a vector
  - Dimension of vector must be equal to number of columns of matrix
    - E.g., mxn matrix multiplied by nx1 vector resulting in mx1 vector
    - Vector is the second operand

Matrix multiplication

- Multiplication of matrix and matrix

$$\begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} * \begin{bmatrix} x & y \\ w & z \end{bmatrix} = \begin{bmatrix} ax + bw & ay + bz \\ cx + dw & cy + dz \\ ex + fw & ey + fz \end{bmatrix}$$

  - Split into multiplications of matrix and vectors and combine the outputs
  - Number of columns of first matrix must be equal to number of rows of second matrix
    - E.g., mxn matrix multiplied by nxp matrix resulting in mxp matrix

Inversion and transposition

- Inversion
  - Inverted matrix of A is denoted by A^{-1}
  - $A*A^{-1} = I$
  - Non-squared matrix does not have inverted matrix
- Transposition
  - Transposed matrix of A is denoted by A^T
  - $A = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}, A^T = \begin{bmatrix} a & c & e \\ b & d & f \end{bmatrix}$
  - $A_{ij} = A^T_{ji}$

Statistics review

- Mean, median
- Variance, standard deviation
- Covariance, correlation
- Law of large numbers

Statistics review

- Mean: arithmetic average of a variable

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

  - Example: Mean of [3, 7, 10] is (3+7+10)/3 = 6.67
- Median: middle value when data is sorted
  - Example #1: Median of [3, 7, 10] is 7
  - Example #2: For [3, 7, 100], the median is still 7

Statistics review

- Variance: average squared deviation from the mean

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

  - What is variance of [2, 4, 6] vs. [1, 9, 10]?
- Standard deviation: square root of variance

Statistics review

- Covariance: measures whether two variables change together

$$\mathrm{Cov}(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_X)(y_i - \mu_Y)$$

- Example

Student	Study Hours (X)	Exam Score (Y)
A	2	50
B	4	60
C	6	70
D	8	80
E	10	90

$$\mu_X = \frac{2+4+6+8+10}{5} = 6$$

$$\mu_Y = \frac{50+60+70+80+90}{5} = 70$$

$$\mathrm{Cov}(X,Y) = \frac{80+20+0+20+80}{5} = \frac{200}{5} = 40$$

Statistics review

- Correlation: measures the strength and direction of a linear relationship of two variables
  - Correlation is covariance that is standardized
- Pearson correlation formula

$$r = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

- Interpretation:
  - r=1: Perfect positive correlation
  - r=−1: Perfect negative correlation
  - r=0: No correlation

Statistics review

- Example

Student	Study Hours (X)	Exam Score (Y)
A	2	50
B	4	60
C	6	70
D	8	80
E	10	90

$$\mu_X = \frac{2+4+6+8+10}{5} = 6$$

$$\mu_Y = \frac{50+60+70+80+90}{5} = 70$$

- Standard deviation of Study Hours (X):
  $\sigma_X \approx 2.83$
- Standard deviation of Exam Scores (Y):
  $\sigma_Y \approx 14.14$

$$\text{Cov}(X,Y) = \frac{80+20+0+20+80}{5} = \frac{200}{5} = 40$$

$$r = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

$$r = \frac{40}{(2.83 \times 14.14)}$$

$$r = \frac{40}{39.99} \approx 1$$

Statistics review

- Law of large numbers: as sample size increases, sample mean approaches population mean
- Example
  - If we flip a fair coin 10 times, heads might appear 3 times (30%).
  - However, after 10,000 flips, the probability will be closer to 50%.