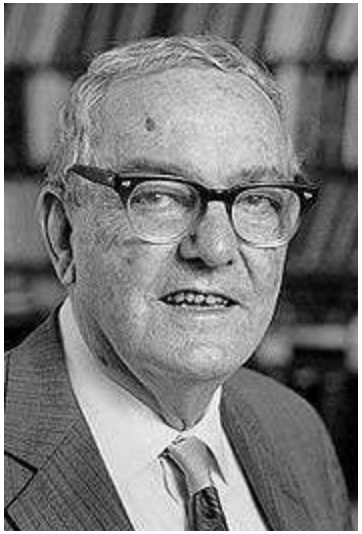
# Applied Artificial Intelligence / Artificial Intelligence

## Machine Learning

# Machine Learning

- **Herbert Alexander Simon**: "Learning is any process by which a system improves performance from experience."
- "Machine Learning is concerned with computer programs that automatically improve their performance through experience. "



**Herbert Simon**
Turing Award 1975
Nobel Prize in Economics 1978

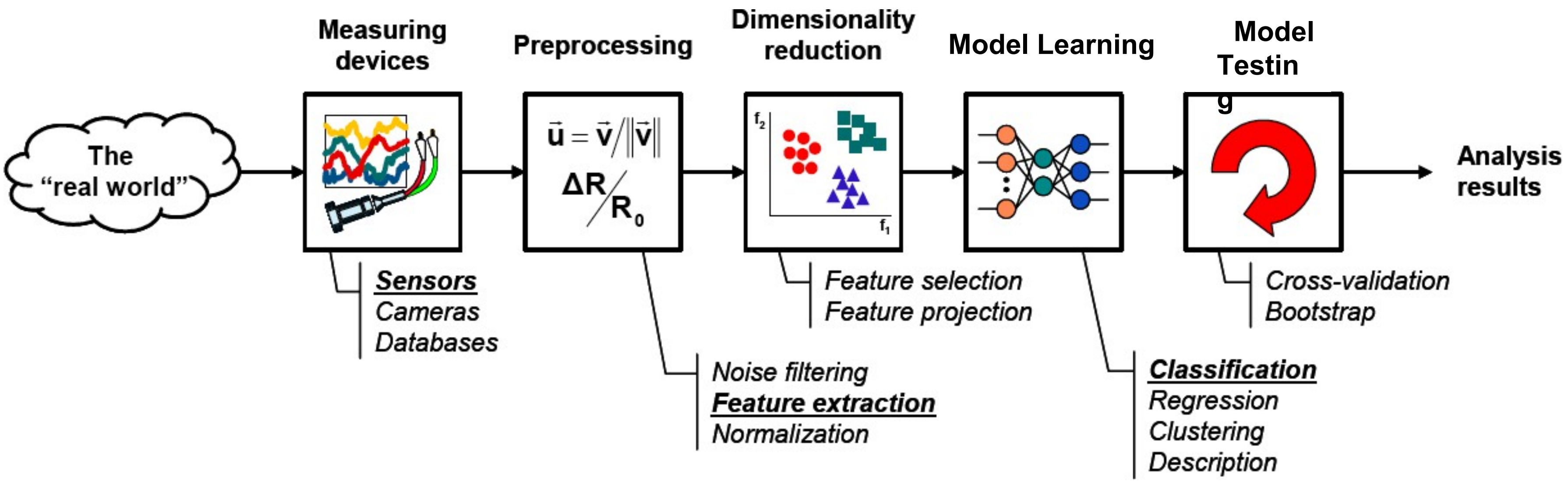
# Why now?

- Flood of available data (especially with the advent of the Internet)
- Increasing computational power
- Growing progress in available algorithms and theory developed by researchers
- Increasing support from industries

# The concept of learning in a ML system

- Learning = Improving with experience at some task
  - Improve over task *T*,
  - With respect to performance measure, *P*
  - Based on experience, *E*.

# The Learning Process



# Machine Learning

- Machine Learning is broadly categorized into:

  - **Supervised Learning** (Labeled Data)

  - **Unsupervised Learning** (Unlabeled Data)

- Used for different types of problems

# Supervised Learning

- Learns from labeled data.
- The model maps inputs to outputs using given examples.
- Goal: Minimize error between predictions and actual labels.
- Common tasks:
  - Classification (Spam detection, Image recognition)
  - Regression (Stock price prediction, House price estimation)

# Unsupervised Learning

- Learns from unlabeled data.
- Finds patterns, structures, or groupings in data.
- Common tasks:
  - Clustering (Customer segmentation, Anomaly detection)
  - Dimensionality Reduction (PCA, t-SNE for visualization)

# Regression vs Classification

| | Regression | Classification |
|---|---|---|
| **Goal** | Predict continuous values | Predict discrete labels |
| **Output Type** | Continuous (Real numbers) | Categorical (Classes/Labels) |
| **Examples** | House price prediction, Temperature forecasting | Spam detection, Disease diagnosis |
| **Algorithms** | Linear Regression, Neural Networks | Logistic Regression, Naïve Bayes, Neural Networks |

# Supervised Learning – Naïve Bayes

| DAY | OUTLOOK | TEMP | HUMIDITY | WIND | PLAY |
|---|---|---|---|---|---|
| D1 | Sunny | Hot | High | Weak | No |
| D2 | Sunny | Hot | High | Strong | No |
| D3 | Overcast | Hot | High | Weak | Yes |
| D4 | Rain | Mild | High | Weak | Yes |
| D5 | Rain | Cool | Normal | Weak | Yes |
| D6 | Rain | Cool | Normal | Strong | No |
| D7 | Overcast | Cool | Normal | Strong | Yes |
| D8 | Sunny | Mild | High | Weak | No |
| D9 | Sunny | Cool | Normal | Weak | Yes |
| D10 | Rain | Mild | Normal | Weak | Yes |
| D11 | Sunny | Mild | Normal | Strong | Yes |
| D12 | Overcast | Mild | High | Strong | Yes |
| D13 | Overcast | Hot | Normal | Weak | Yes |
| D14 | Rain | Mild | High | Strong | No |

# Naïve Bayes Classification

**Step#1:** Total Rows = 14

| | |
|---|---|
| Total YES = 9 | Total NO = 5 |
| Probability (YES) = 9/14 | Probability (NO) = 5/14 |

**Step#2:** Make a table for every attribute. (4 attributes, 4 tables)

| **Attribute: Outlook** | | |
|---|---|---|
| | **YES** | **NO** |
| **Sunny** | 2/9 | 3/5 |
| **Overcast** | 4/9 | 0/5 |
| **Rain** | 3/9 | 2/5 |
| **Total** | **9/9** | **5/5** |

| **Attribute: Temperature** | | |
|---|---|---|
| | **YES** | **NO** |
| **Hot** | 2/9 | 2/5 |
| **Mild** | 4/9 | 2/5 |
| **Cool** | 3/9 | 1/5 |
| **Total** | **9/9** | **5/5** |

| **Attribute: Humidity** | | |
|---|---|---|
| | **YES** | **NO** |
| **High** | 3/9 | 4/5 |
| **Normal** | 6/9 | 1/5 |
| **Total** | **9/9** | **5/5** |

| **Attribute: Wind** | | |
|---|---|---|
| | **YES** | **NO** |
| **Strong** | 3/9 | 3/5 |
| **Weak** | 6/9 | 2/5 |
| **Total** | **9/9** | **5/5** |

# Naïve Bayes Classification

- Predict(X), X= Sunny, Cool, High, Strong
  - Take positive for all and multiply them
  - Take negative for all and multiply them
  - Compare both, the greater value is the prediction
  - E.g.,
    - $\text{Probability (YES for X)} = \frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14} = 0.005$
    - $\text{Probability (NO for X)} = \frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5} \times \frac{5}{14} = 0.02$
    - $0.02 > 0.005,$ so the label is NO.

# Unsupervised Learning

K-means Clustering

# K-means Clustering

**Input:**
- A dataset *D* with *n* data points
- Number of clusters *k*
- A distance metric (e.g., Euclidean distance)
- A stopping criterion (e.g., convergence of centroids or max iterations)

**Output:**
- *k* cluster centroids
- Cluster assignments for each data point

1. **Initialize centroids**:
   Select k initial cluster centroids randomly from the dataset or using a specific method.
2. **Repeat until convergence**:
   a. **Assign each data point to the nearest centroid**:
      For each data point $x_i \in D$, compute the distance to all centroids.
      Assign $x_i$ to the cluster with the closest centroid.
   b. **Update centroids**:
      For each cluster $j$, recompute the centroid as the mean of all points assigned to that cluster.
   c. **Check stopping criterion**:
      If centroids do not change significantly or max iterations reached, stop.
3. **Return final cluster centroids and assignments**.

# K-means Clustering - Example

**Centroids**
**$C_1$:** (2, 10)
**$C_2$:** (5, 8)
**$C_3$:** (1, 2)

| Datapoints | | | Distance to Centroid | | | Cluster |
|---|---|---|---|---|---|---|
| | | | $C_1$ | $C_2$ | $C_3$ | |
| D1 | 2 | 10 | | | | |
| D2 | 2 | 5 | | | | |
| D3 | 8 | 4 | | | | |
| D4 | 5 | 8 | | | | |
| D5 | 7 | 5 | | | | |
| D6 | 6 | 4 | | | | |
| D7 | 1 | 2 | | | | |
| D8 | 4 | 9 | | | | |

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

# K-means Clustering - Example

**Centroids**
**$C_1$:** (2, 10)
**$C_2$:** (5, 8)
**$C_3$:** (1, 2)

**New Centroids**
**$C_1$:** (2, 10)
**$C_2$:** (6, 6)
**$C_3$:** (1.5, 3.5)

| Datapoints | | | Distance to Centroid | | | Cluster |
|---|---|---|---|---|---|---|
| | | | $C_1$ | $C_2$ | $C_3$ | |
| D1 | 2 | 10 | 0 | 3.61 | 8.06 | 1 |
| D2 | 2 | 5 | 5 | 4.24 | 3.16 | 3 |
| D3 | 8 | 4 | 8.49 | 5 | 7.28 | 2 |
| D4 | 5 | 8 | 3.61 | 0 | 7.21 | 2 |
| D5 | 7 | 5 | 7.07 | 3.61 | 6.71 | 2 |
| D6 | 6 | 4 | 7.21 | 4.12 | 5.39 | 2 |
| D7 | 1 | 2 | 8.06 | 7.21 | 0 | 3 |
| D8 | 4 | 9 | 2.24 | 1.41 | 7.62 | 2 |

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

# K-means Clustering - Example

**Centroids**
**$C_1$:** (2, 10)
**$C_2$:** (6, 6)
**$C_3$:** (1.5, 3.5)

**New Centroids**
**$C_1$:** (3, 9.5)
**$C_2$:** (6.5, 5.25)
**$C_3$:** (1.5, 3.5)

| Datapoints | | | Distance to Centroid | | | Old Cluster | New Cluster |
|---|---|---|---|---|---|---|---|
| | | | $C_1$ | $C_2$ | $C_3$ | | |
| D1 | 2 | 10 | 0 | 5.66 | 6.52 | 1 | 1 |
| D2 | 2 | 5 | 5 | 4.12 | 1.58 | 3 | 3 |
| D3 | 8 | 4 | 8.49 | 2.83 | 6.52 | 2 | 2 |
| D4 | 5 | 8 | 3.61 | 2.24 | 5.70 | 2 | 2 |
| D5 | 7 | 5 | 7.07 | 1.41 | 5.70 | 2 | 2 |
| D6 | 6 | 4 | 7.21 | 2 | 4.53 | 2 | 2 |
| D7 | 1 | 2 | 8.06 | 6.40 | 1.58 | 3 | 3 |
| D8 | 4 | 9 | 2.24 | 3.61 | 6.04 | 2 | 1 |

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

# K-means Clustering - Example

**Centroids**
**$C_1$**: (3, 9.5)
**$C_2$**: (6.5, 5.25)
**$C_3$**: (1.5, 3.5)

**New Centroids**
**$C_1$**: (3.67, 9)
**$C_2$**: (7, 4.33)
**$C_3$**: (1.5, 3.5)

| Datapoints | | | Distance to Centroid | | | Old Cluster | New Cluster |
|---|---|---|---|---|---|---|---|
| | | | $C_1$ | $C_2$ | $C_3$ | | |
| D1 | 2 | 10 | 1.12 | 6.54 | 6.52 | 1 | 1 |
| D2 | 2 | 5 | 4.61 | 4.51 | 1.58 | 3 | 3 |
| D3 | 8 | 4 | 7.43 | 1.95 | 6.52 | 2 | 2 |
| D4 | 5 | 8 | 2.50 | 3.13 | 5.70 | 2 | 1 |
| D5 | 7 | 5 | 6.02 | 0.56 | 5.70 | 2 | 2 |
| D6 | 6 | 4 | 6.26 | 1.35 | 4.53 | 2 | 2 |
| D7 | 1 | 2 | 7.76 | 6.39 | 1.58 | 3 | 3 |
| D8 | 4 | 9 | 1.12 | 4.51 | 6.04 | 1 | 1 |

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

# K-means Clustering - Example

**Centroids**
**$C_1$**: (3.67, 9)
**$C_2$**: (7, 4.33)
**$C_3$**: (1.5, 3.5)

| Datapoints | | | Distance to Centroid | | | Old Cluster | New Cluster |
|---|---|---|---|---|---|---|---|
| | | | $C_1$ | $C_2$ | $C_3$ | | |
| D1 | 2 | 10 | 1.94 | 7.56 | 6.52 | 1 | 1 |
| D2 | 2 | 5 | 4.33 | 5.04 | 1.58 | 3 | 3 |
| D3 | 8 | 4 | 6.62 | 1.05 | 6.52 | 2 | 2 |
| D4 | 5 | 8 | 1.67 | 4.18 | 5.70 | 1 | 1 |
| D5 | 7 | 5 | 5.21 | 0.67 | 5.70 | 2 | 2 |
| D6 | 6 | 4 | 5.52 | 1.05 | 4.53 | 2 | 2 |
| D7 | 1 | 2 | 7.49 | 6.44 | 1.58 | 3 | 3 |
| D8 | 4 | 9 | 0.33 | 5.55 | 6.04 | 1 | 1 |

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$