

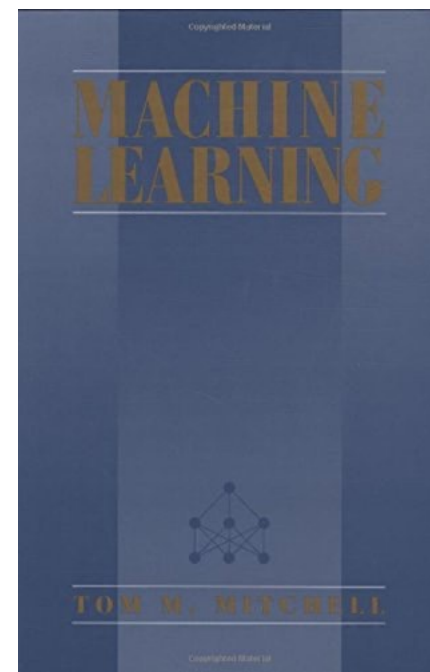
# Definition

---

Adapted from Mitchell, *Machine Learning*, 1997.

## **A definition of learning from data:**

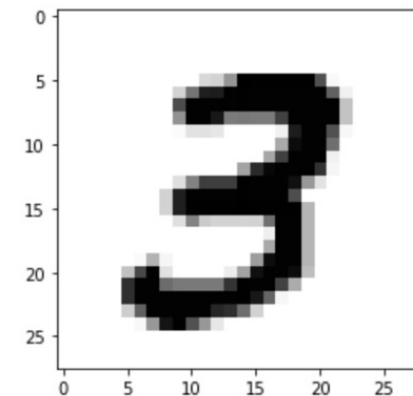
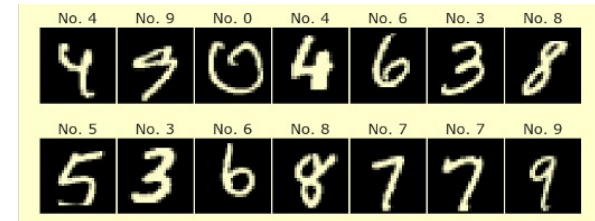
Consider a collection of tasks  $T$ , a performance measure  $P$ , a baseline strategy  $B$ , and an algorithm  $A$  which depends on a set of training data  $D$ . The algorithm  $A$  is said to **learn** from the data  $D$ , if its performance at tasks in  $T$ , as measured by  $P$ , is better than the baseline strategy  $B$ .



# Example

---

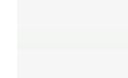
- **Task:** Identify the values of handwritten digits (given pixel values).
- **Baseline strategy:** Always predict 0.
- **Training data:** Handwritten digits together with their correct values.
- **Performance measure:** Percentage of digits correctly identified.



# Example

- **Task:** Determine if an email is spam.
- **Baseline strategy:** Predict spam if the email contains 3 or more exclamation points (!).
- **Training data:** Emails that have been identified as spam/not spam.
- **Performance measure:** Percentage correctly identified.

*Notification: We Have A Surprise For You!*



## CASH APP

***You've Been selected this week as a winner!***

***Win \$1000 To Your CashApp***

*You Win a \$1000 Cash App*

*if you're receiving this email Today, click to the attached site and won \$1000.00!*

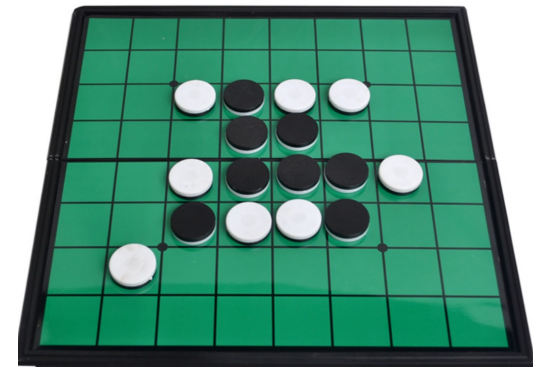
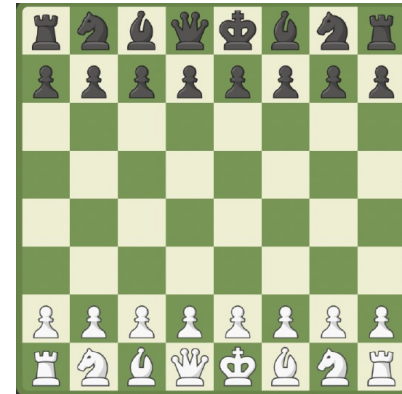
***Congratulations you qualified!***

Enter Your Details on the next page and  
*Click The "ACCEPT" Button To Continue...*

**ACCEPT**

# Example

- **Task:** Assign a value to a given board state.
- **Baseline strategy:** 100 if white has won, -100 if black has won, 0 otherwise
- **Training data:** Move history for a collection of games.
- **Performance measure:** Percentage of games won when choosing the move that optimizes board value.



# Example

---

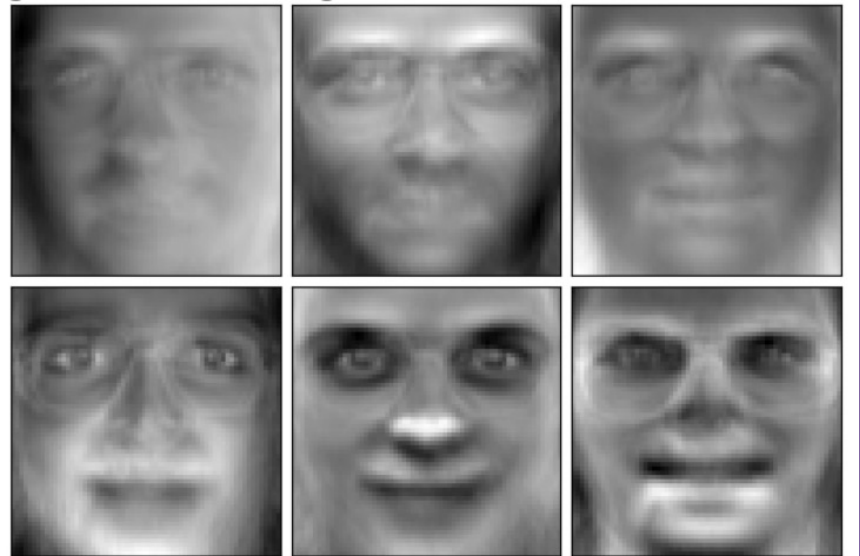
- **Task:** Generate art.
- **Baseline strategy:** Random pixel values.
- **Training data:** Collection of artworks together with evaluations by an expert.
- **Performance measure:** Was an expert tricked into thinking the artwork was made by a human?



# Example

---

- **Task:** Reduce the dimensionality of images of faces.
- **Baseline strategy:** Keep only the center-most 36 pixel values
- **Training data:** A collection of images of faces.
- **Performance measure:** Similarity of the reduced face to the original image.





# Example

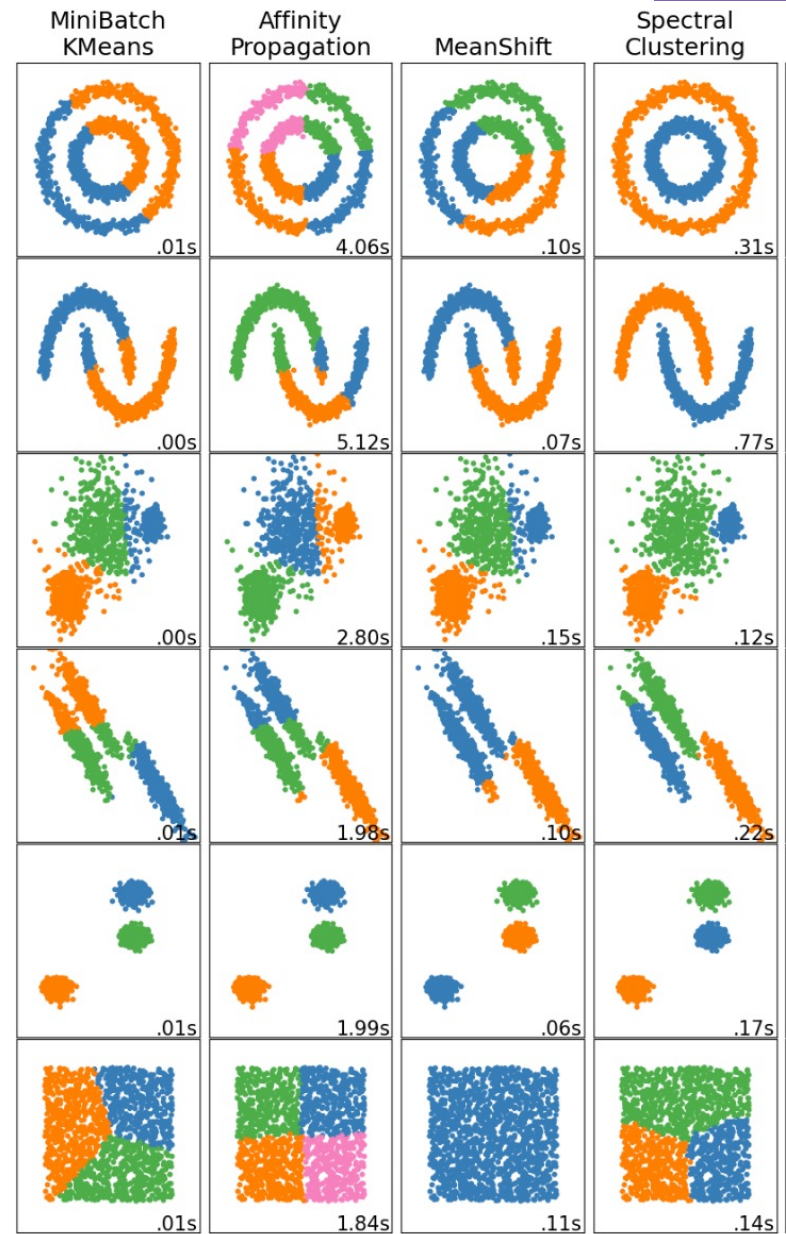
---

- **Task:** Estimate the probability that a Titanic passenger survived.
- **Baseline strategy:** Use the average survival rate as prediction.
- **Training data:** Survival outcomes and passenger characteristics.
- **Performance measure:** Log loss



# Example

- **Task:** Identify  $K$  distinct centroid locations.
- **Baseline strategy:** Apply K-Means clustering to the first 10 sample points.
- **Training data:** 100 sample points.
- **Performance measure:** Average distance of a sample point to the nearest centroid.





# Example

---

- **Task:** Predict prices of houses in King County, Washington, based on characteristics.
- **Baseline strategy:** Always predict the median house value.
- **Training data:** Prices of certain houses in the county.
- **Performance measure:** Mean absolute error.



Source: Wikimedia Commons, Hannah Lewis House, Jon Roanhaus