# Machine Learning (Introduction)

Prof K R Chowdhary

MBM University

September 5, 2024

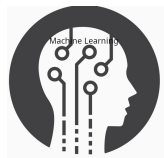
# Introduction to ML

- Machine Learning (ML) means we wish to program the computers so that they can "learn" from input available to them.
- The input to a learning algorithm is *training data* that represents experience, and the output is some expertise, which usually takes the form of another computer program that can perform some task.
- For the formal and mathematical understanding of this concept, we should be explicit about what we mean by following terms:
  1. What is the training data that our programs uses?
  2. How can the process of learning be automated?
  3. How can we evaluate the success of such a process (i.e., the quality of the output of a learning program)?

# Machined Learning Defined

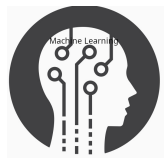
**Example 1:** *Rats' Learning to avoid poisonous baits*:

- When rats encounter food items with novel look or smell, they will first eat very small amounts of it, and subsequent feeding will depend on the flavor of the food and its physiological effect.
- If the food produces an ill effect, the novel food will often be associated with the illness, and subsequently, the rats will not eat it.
- Clearly, there is a learning mechanism in play here – the rat used past experience with some food to acquire expertise in detecting the safety of this food.
- If past experience with the food was negatively labeled, the rat predicts that it will also have a negative effect when encountered in the future.



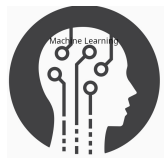
# Application in Spam Filtering

- Inspired by above example of successful learning, let us demonstrate a typical machine learning task. Suppose we would like to program a machine that learns how to filter *spam e-mails*.
- A naive solution would be similar to the way rats learn how to avoid poisonous baits. The machine will simply memorize all previous e-mails that had been labeled as spam e-mails by the human user.
- When a new e-mail arrives, the machine will search for it in the set of previous spam e-mails. If it matches one of them, it will be moved into trash folder, else, to the user's inbox.



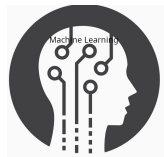
# Problem with 'bait's shyness Learning

- It uses "learning by memorization" approach, hence lacks the ability to label new e-mail messages.
- A successful learner should progress from individual examples to broader generalization (inductive reasoning).
- In "bait's shyness", having seen an example of a certain type of food, rats apply their attitude toward new, unseen examples of food, of similar smell and taste.
- To achieve generalization in the spam filtering, the learner can scan the previously seen e-mails, and extract a set of words whose appearance is indicative of spam.
- When a new e-mail arrives, the machine can check whether one of the suspicious words appears in it, and labels it accordingly.
- Such a system would potentially be able to correctly predict the label of unseen e-mails.



# Criterias about the Theory of ML

- Incorporation of *prior knowledge*, biasing the learning process, is essential for the success of learning algorithms.
- Following are the central criteria about the theory of machine learning:
  1. development of tools for expressing domain expertise,
  2. translating it into a learning bias, and
  3. quantifying the effect of such a bias on the success of learning.
- The stronger the prior knowledge, the easier it is to learn from further examples.
- Also, stronger these prior assumptions are, the less flexible the learning is.



# Justification for Machine Learning

- In the early days of "intelligent" applications, many systems used hand-coded rules of "if" and "else" decisions to process data or adjust to user input.
- In spam filtering, one could make up a blacklist of words that would result in an email being marked as spam.
- Manually crafting decision rules is feasible for some applications, those in which humans have a good understanding of the process to model.
- However, using hand coded rules to make decisions has two major disadvantages:
  - The logic required to make a decision is specific to a single domain and task. Changing the task even slightly might require a rewrite of the whole system.
  - Designing rules requires a deep understanding of how a decision should be made by a human expert.

# Why hand-coded System are not sufficient?

- One example where this hand-coded approach will fail is in detecting faces in images.
- Today, every smartphone can detect face in an image, often covered in boxes while taking snaps. However, face detection was an unsolved problem until as recently as 2001.
- The main problem is that the way in which pixels (which make up an image in a computer) are "perceived" by the computer is very different from how humans perceive a face.
- This difference in representation makes it basically impossible for a human to come up with a good set of rules to describe what constitutes a face in a digital image.
- Using machine learning, however, simply presenting a program with a large collection of images of faces is enough for an algorithm to determine what characteristics are needed to identify a face.