Applied Machine Learning

Dr. Mohsin Kamal

Department of Electrical Engineering
National University of Computer and Emerging Sciences, Lahore, Pakistan

Figure 1 : Photo tagging using ML

Figure 2 : Recognition of spam emails

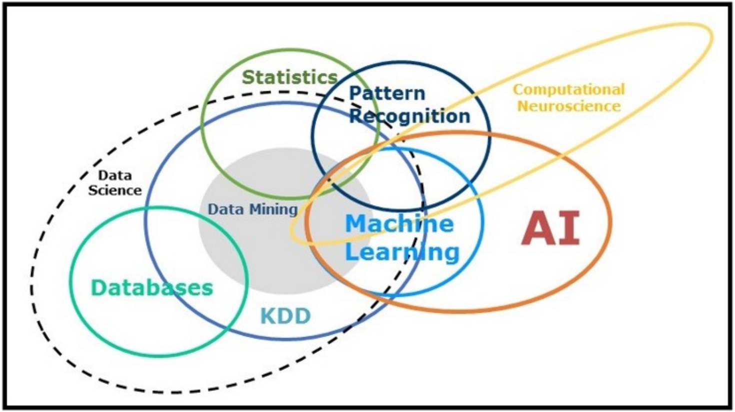


Figure 3 : Overlapping of ML with other progressive fields

Why is Machine Learning so prevalent today?

- It has grown out of the field from Artificial Intelligence (AI)
- Developed as new capability for computers

Examples:

- Database mining (turning raw data into useful information)
  - Large datasets from growth of automation/web.
  - E.g., Web click data, medical records, biology, engineering
- Applications which were not possible to be programmed by hand.
  - E.g., Autonomous vehicles, handwriting recognition, Computer Vision.
- Self-customizing program
  - E.g., Amazon, Netflix product recommendations
- Understanding human learning (brain, real AI).

Figure 4 : TESLA's self driving car.

Definition

Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.
Arthur Samuel (1959)

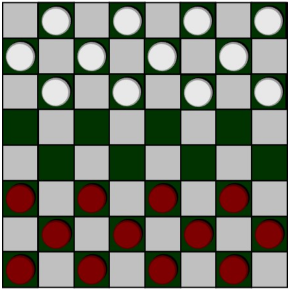


Figure 5 : Checker/Draughts game.

Definition

Tom Mitchell (1998) Well-posed Learning Problem: A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

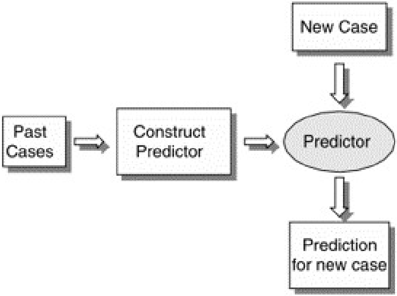
Definition

Modern Definition of ML:

A subset of artificial intelligence (AI), machine learning (ML) is the area of computational science that focuses on analyzing and interpreting patterns and structures in data to enable learning, reasoning, and decision making outside of human interaction.

Machine Learning (Prediction or Forecasting)

- Learning from samples of past experience and projecting on new examples
- The future has to be similar to the past
- Inductive learning approach
- Dominant approach in machine learning



Learning task

Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task T in this setting?

Options:

1. Classifying emails as spam or not spam.
2. Watching you label emails as spam or not spam.
3. The number (or fraction) of emails correctly classified as spam/not spam.
4. None of the above-this is not a machine learning problem.

Learning task

Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task T in this setting?

Options:

1. Classifying emails as spam or not spam.
2. Watching you label emails as spam or not spam.
3. The number (or fraction) of emails correctly classified as spam/not spam.
4. None of the above-this is not a machine learning problem.

Learning task

Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the Experience E in this setting?
Options:

1. Classifying emails as spam or not spam.
2. Watching you label emails as spam or not spam.
3. The number (or fraction) of emails correctly classified as spam/not spam.
4. None of the above-this is not a machine learning problem.

Learning task

Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the performance measure P in this setting?

Options:

1. Classifying emails as spam or not spam.
2. Watching you label emails as spam or not spam.
3. The number (or fraction) of emails correctly classified as spam/not spam.
4. None of the above-this is not a machine learning problem.

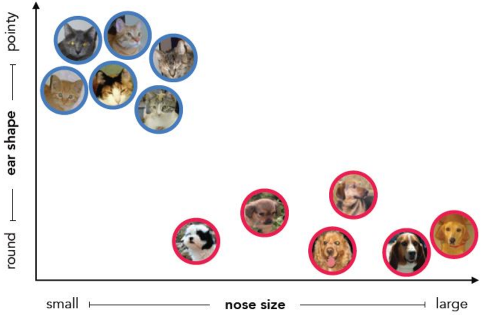
Data collection

- A computer must be trained to recognize the difference between these two types of animals by learning from a batch of examples, typically referred to as a **training set** of data
- The larger and more diverse the training set the better a computer (or human) can perform a learning task, since exposure to a wider breadth of examples gives the learner more experience.

Distinguishing Cats from Dogs: a Machine Learning Approach

Feature design

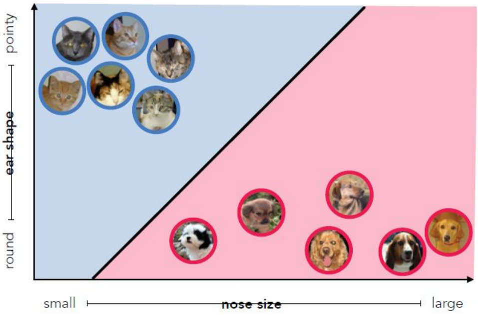
- In order to successfully train a computer to perform this task (and any machine learning task more generally) we need to provide it with properly designed features or, ideally, have it find or learn such features itself.
- Suppose we can easily extract the following two features from each image in the training set: size of nose relative to the size of the head, ranging from small to large, and shape of ears, ranging from round to pointy



Distinguishing Cats from Dogs: a Machine Learning Approach

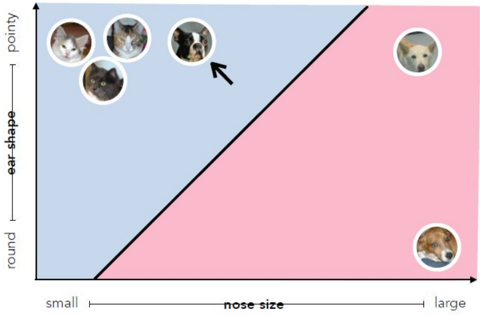
Model training

- The machine should find a line or a curve that separates the cats from the dogs in our carefully designed feature space.
- We must find the right values for its two parameters - a slope and vertical intercept.
- The process of determining proper parameters relies on a set of tools known as **mathematical optimization**.



Model validation

- To validate the efficacy of our trained learner we now show the computer a batch of previously unseen images of cats and dogs, referred to generally as a **validation set** of data, and see how well it can identify the animal in each image.
- The misidentification of the single dog (a Boston terrier) is largely the result of our choice of features, which we designed based on the training set and to some extent our decision to use a linear model (instead of a nonlinear one).



Distinguishing Cats from Dogs: a Machine Learning Approach

To summarize:

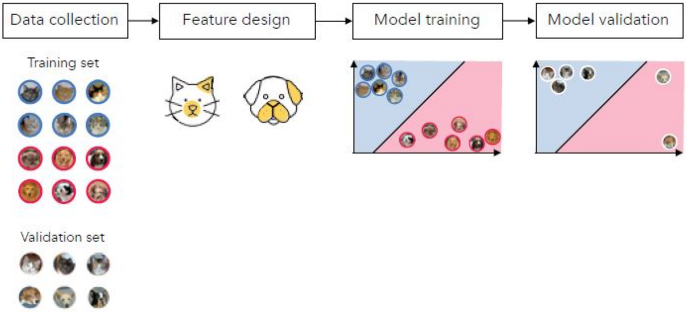


Figure 6 : The schematic pipeline of our toy cat-versus-dog classification problem. The same general pipeline is used for essentially all machine learning problems.

Distinguishing Cats from Dogs: a Machine Learning Approach

Revisit Machine Learning

- Minimum effort
- Minimum domain expertise (subjective)
- Data is available in abundance anyways
- Computational resources are cheap and available as service through Cloud

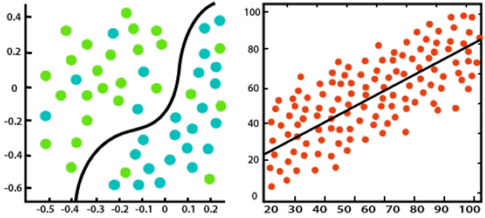
Machine Learning algorithms

- Supervised learning
  - Regression
  - Classification
- Unsupervised learning
  - Clustering
  - Auto-encoders
  - Dimensionality reduction
  - Topic modeling
- Other: Reinforcement learning, recommender systems.

Supervised Machine Learning

Applicable to a wide array of situations and data types, Supervised Learning Problems come in two forms:

- Regression
- Classification



Classification Regression

Housing price prediction

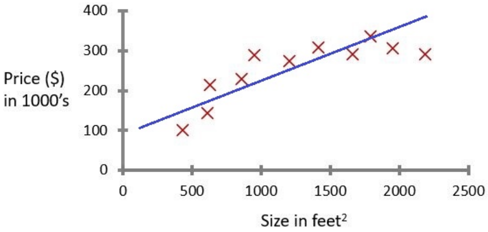


Figure 7 : Price in $ vs. area of house

Supervised Learning: "Right Answers" Given	Regression: Predict continuous valued output (price)

Supervised Machine Learning

Cancer (malignant, benign)

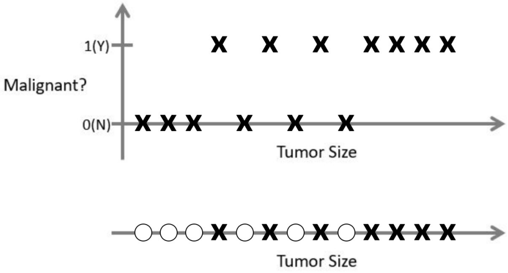


Figure 8 : Discussion on the tumor

Classification

Discrete valued output (0 or 1)

Supervised Machine Learning

Cancer (malignant, benign)

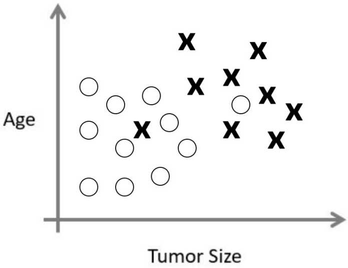


Figure 9 : Discussion on the tumor

Other features could be: Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape

Learning task

You are running a company, and you want to develop learning algorithms to address each of the two problems.
Problem 1: You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.
Problem 2: You would like software to examine individual customer accounts, and for each account decide if it has been hacked/compromised.
Should you treat these as classification or as regression problems?

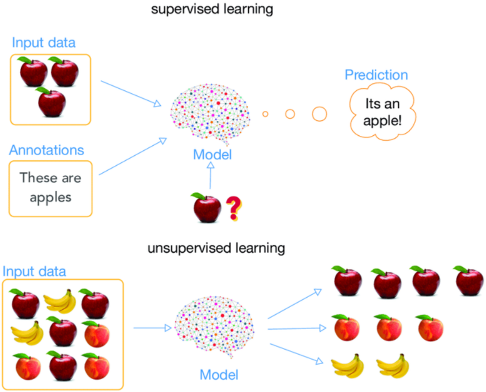
1. Treat both as classification problems.
2. Treat problem 1 as a classification problem, problem 2 as a regression problem.
3. Treat problem 1 as a regression problem, problem 2 as a classification problem.
4. Treat both as regression problems.

Learning task

You are running a company, and you want to develop learning algorithms to address each of the two problems.
Problem 1: You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.
Problem 2: You would like software to examine individual customer accounts, and for each account decide if it has been hacked/compromised.
Should you treat these as classification or as regression problems?

1. Treat both as classification problems.
2. Treat problem 1 as a classification problem, problem 2 as a regression problem.
3. Treat problem 1 as a regression problem, problem 2 as a classification problem.
4. Treat both as regression problems.

Supervised and unsupervised learning



Unsupervised Learning

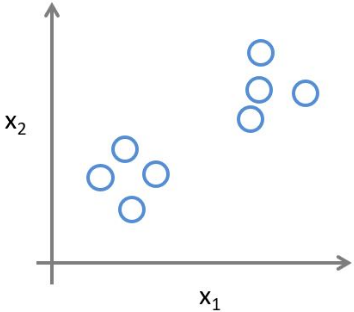


Figure 10 : Example of unsupervised learning (no labels)

Separating data into clusters is called clustering algorithm

Clustering algorithm

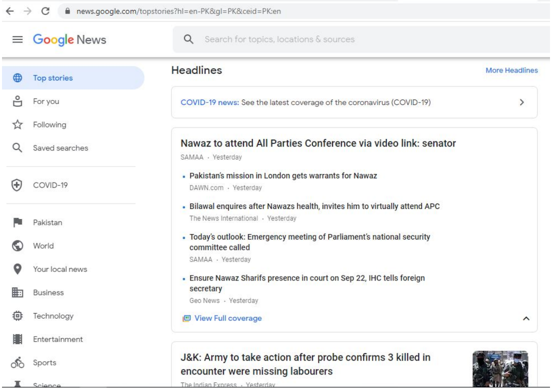


Figure 11 : Example of clustering algorithm

Clustering algorithm

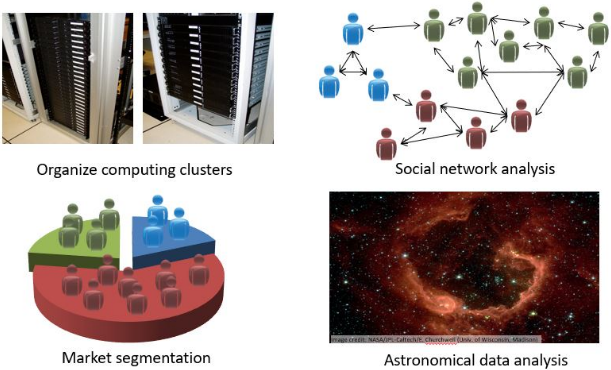


Figure 12 : Example of clustering algorithm

Unsupervised Learning

Dimensionality Reduction

- There may be a high-dimensional feature set with many irrelevant features.
- Reduce the problem space by
  - Extracting new features.
  - Retaining only those features that have better discriminative ability.

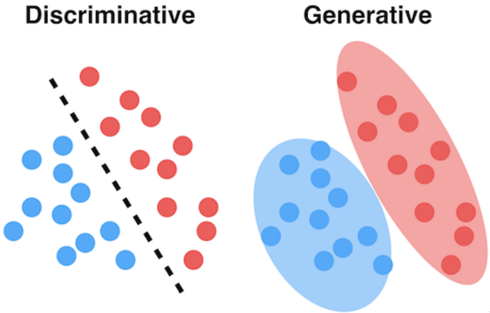
Curse of Dimensionality

- Too many features
- Many features have little contribution towards prediction
- Not enough instances considering the given feature set
- To train a robust model we need few features and many samples
- **At what ratio!** Depends on the nature of prediction model and application

Discriminative vs Generative Models

Discriminative models also referred to as conditional models or backward models as they estimate a decision boundary by using the knowledge inferred from the observed data.

Generative models predict the features of a model that would have generated the data of each time. Such models can be used to generate more such data.



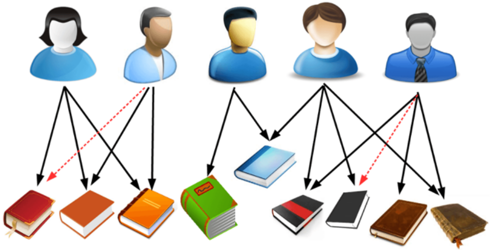
Learning task

Of the following examples, which would you address using an unsupervised learning algorithm? (Check all that apply.)

1. Given email labeled as spam/not spam, learn a spam filter.
2. Given a set of news articles found on the web, group them into sets of articles about the same stories.
3. Given a database of customer data, automatically discover market segments and group customers into different market segments.
4. Given a dataset of patients diagnosed as either having diabetes or not, learn to classify new patients as having diabetes or not.

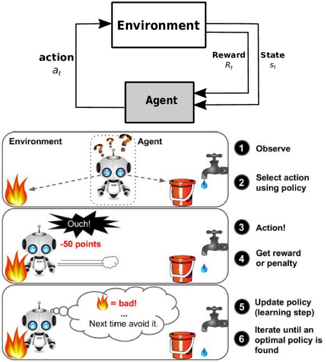
Recommender systems

A recommendation system is an artificial intelligence or AI algorithm, usually associated with machine learning, that uses Big Data to suggest or recommend additional products to consumers.



Reinforcement learning

- The learning system, called an agent in this context, can observe the environment, select and perform actions, and get rewards in return (or penalties in the form of negative rewards)
- It must then learn by itself what is the best strategy, called a policy, to get the most reward over time. A policy defines what action the agent should choose when it is in a given situation.



Hypothesis: A proposed solution to the problem. Generally random within the given constraints.
Objective Function: A function that gives an estimate of how far the hypothesis (proposed solution) is from the actual one by comparing outputs.

Optimizing Objective Function

- Minimizing or maximizing the objective function to reach at optimum solution.
- If the objective function represents profit, the optimization process will attempt to maximize it.
- However, if the objective function represent costs, errors, frauds then the optimization process will attempt to minimize it.

- We will representing the objective function as a cost function i.e., how many and how much mistakes the model is making in predicting.
- So we will be more often minimizing the cost function in optimization
- Therefore, we may also refer to our optimization function as cost function.

Experience

- The model is going to learn based on what is experienced (available data).
- If the data is not a good representation of the real-world, the model is going to learn the same way.

Evaluation

- The evaluation criteria that we are using for deciding if a model is doing good.
- In appropriate evaluation techniques can also take the learning astray.
- Existing problems have these aspects decided but a new problem requires them to be decided carefully.