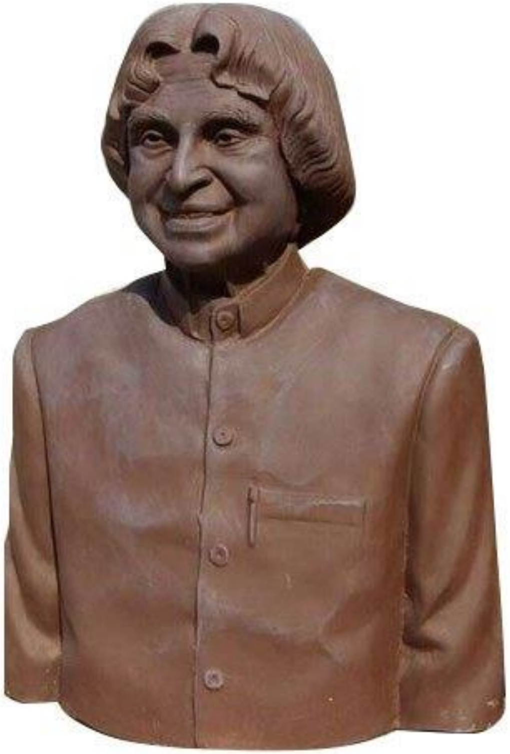
Introduction to Machine Learning and Its Applications

Learning?

Can you recognize these images?



How do you recognize it?

Origin of Machine Learning?

…...Lies in very early efforts of understanding Intelligence.

What is Intelligence?

- It can be defined as the ability to comprehend; to understand and profit from experience.
- Capability to acquire and Apply Knowledge.

What is Machine Learning?

- It is very hard to write programs that solves problems like recognizing handwritten documents, recognizing a 3D object from a novel viewpoint in new lighting conditions in a cluttered scene.
  - We don’t know what program to write because we don’t know how its done in our brain.
  - Even if we had a good idea about how to do it, the program might be extremely complicated.

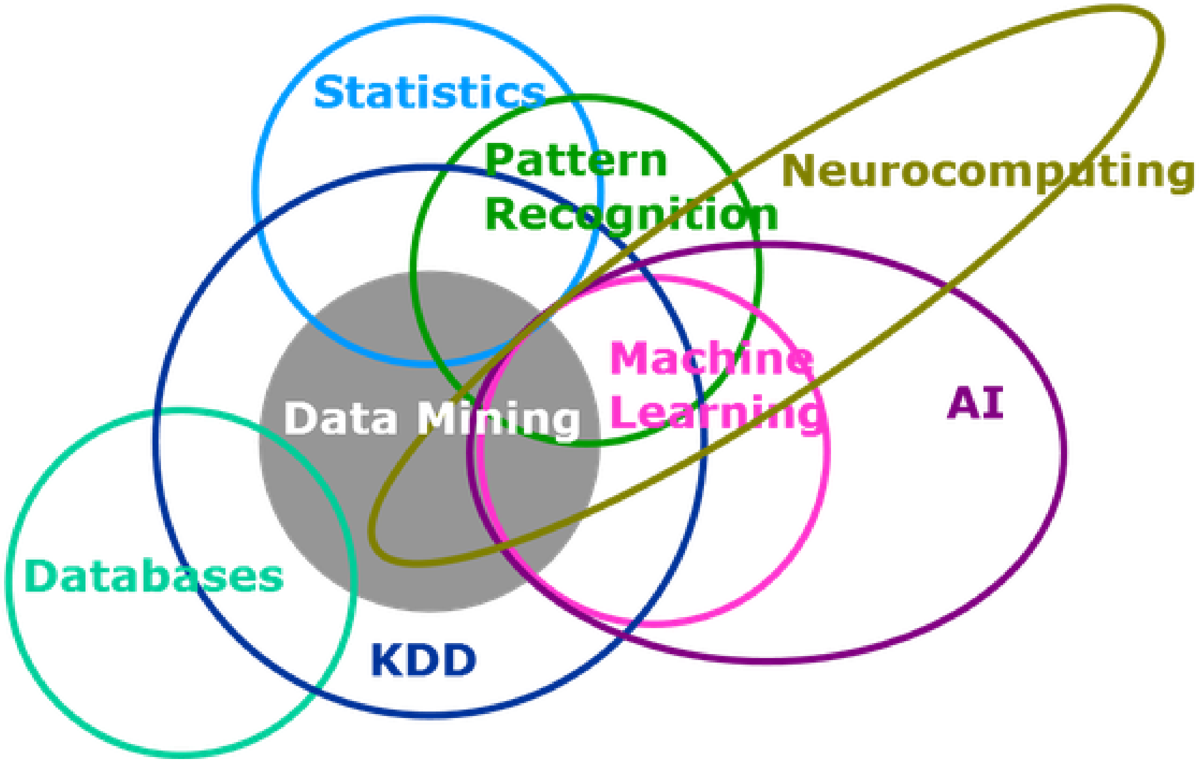
Machine Learning Approach

- Instead of writing a program by hand for each specific task, we collect lots of examples that specify the correct output for a given input.
- A machine learning algorithm then takes these examples and produces a program that does the job.
  - The program produced by the learning algorithm may look very different from a typical hand-written program. It may contain millions of numbers.
  - If you do it right, the program works for new cases as well as the ones we trained it on.
  - If the data changes, the program can change too by training on the new data.
- Massive amounts of computation are now cheaper than paying someone to write a task-specific program.

Machine Learning

"Field of study that gives computers the ability to learn without being explicitly programmed".

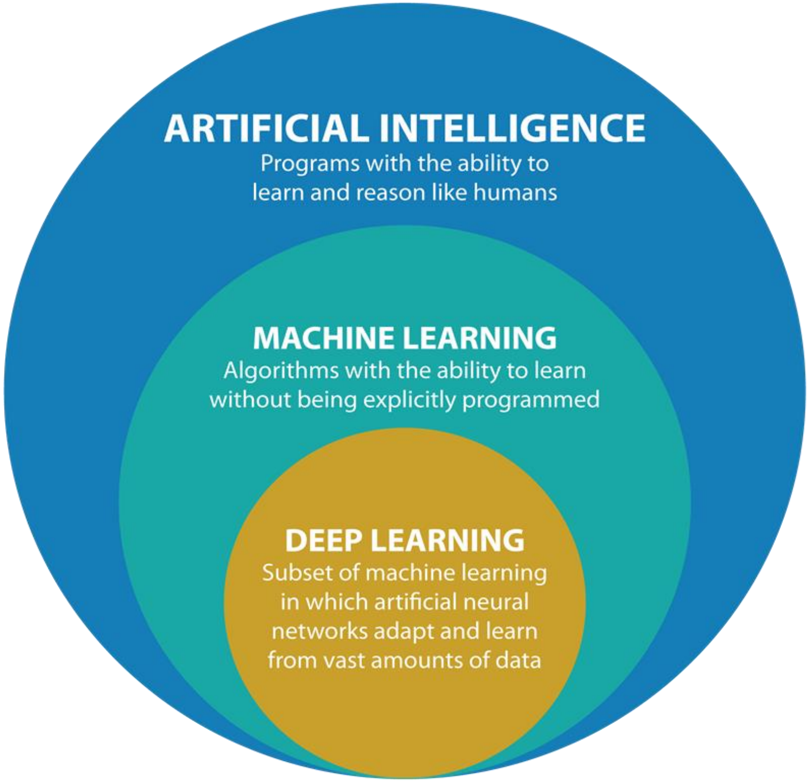
Arthur Samuel, 1959



Machine Learning

"**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".**

Tom Mitchell (1998)



Machine Learning

“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".

Example:

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?

- Classifying emails as spam or not spam. (Task T)
- Watching you label emails as spam or not spam. (Experience E)
- The number (or fraction) of emails correctly classified as spam/not spam. (Performance measure P)

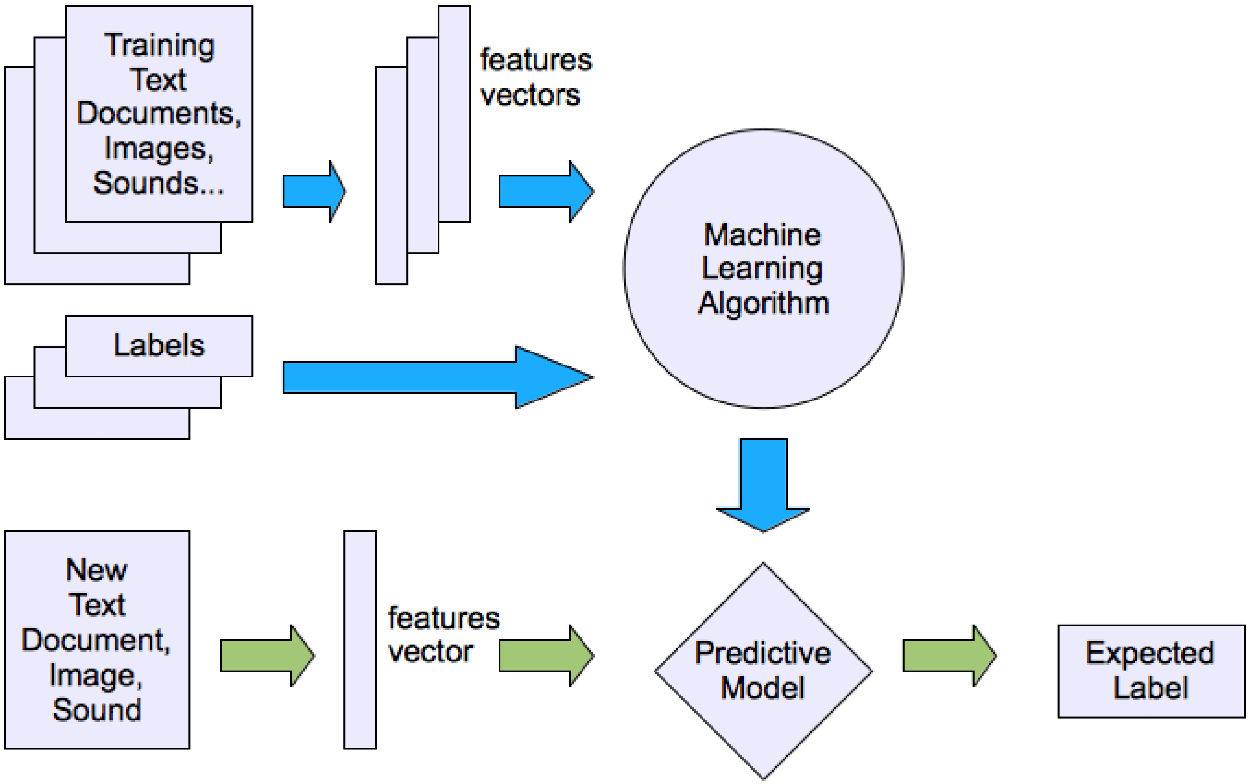
Learning Algorithms

- **Supervised learning** ($\{x_n \in R^d, y_n \in R\}_{n=1}^N$)
  - Prediction
  - Classification (discrete labels), Regression (real values)
- **Unsupervised learning** ($\{x_n \in R^d\}_{n=1}^N$)
  - Clustering
  - Probability distribution estimation
  - Finding association (in features)
  - Dimension reduction (not all)
- **Semi-supervised learning**
- **Reinforcement learning**

Learning Algorithms

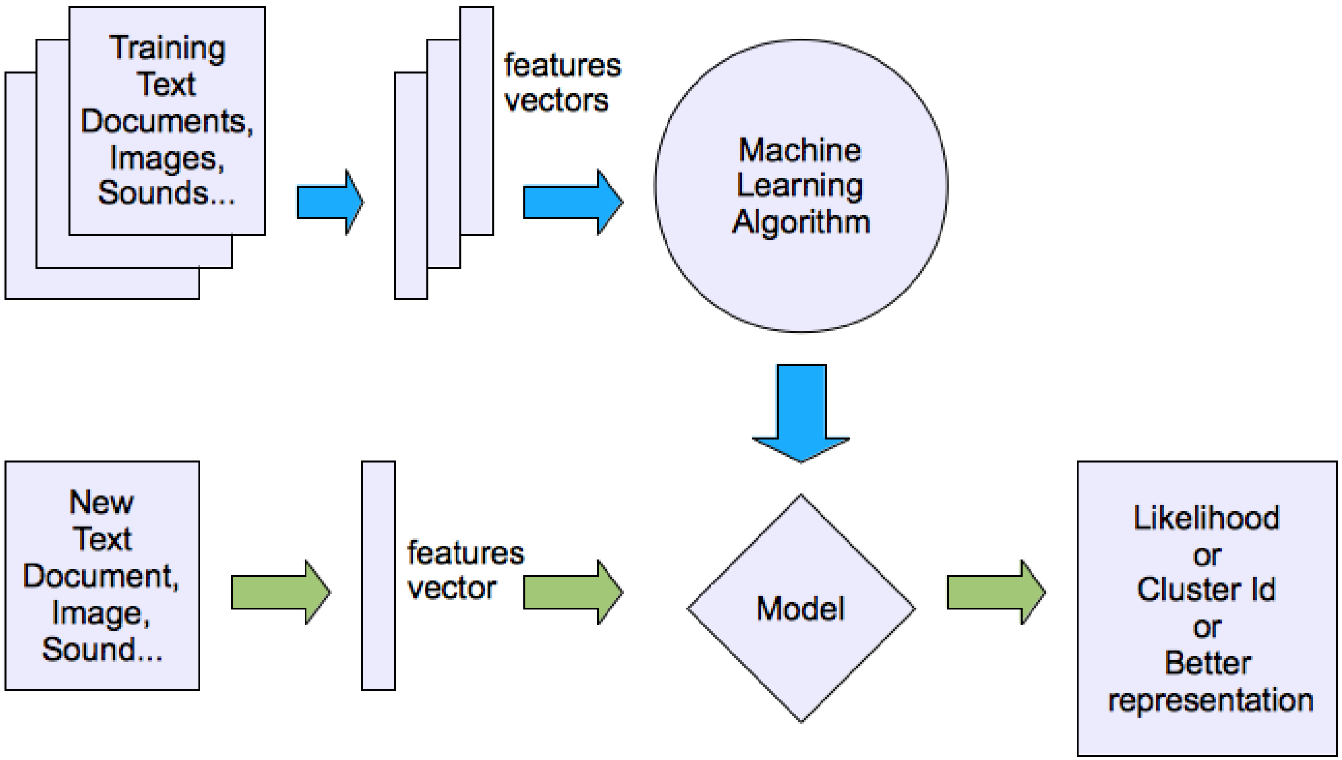
Machine learning structure

- Supervised learning



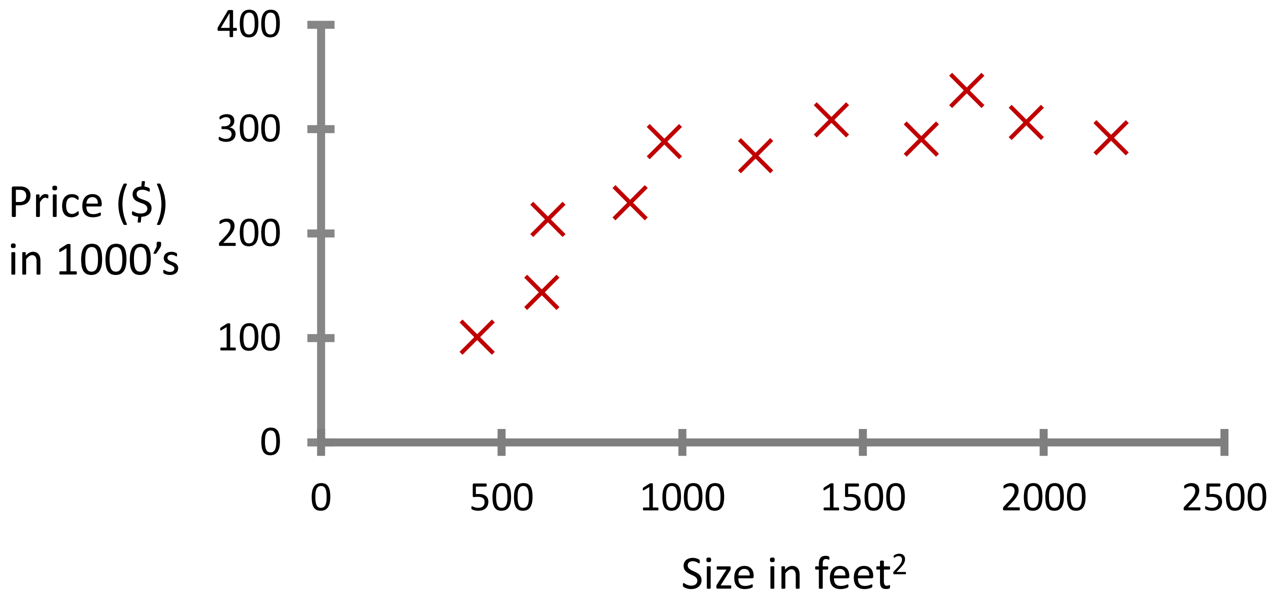
Machine learning structure

- Unsupervised learning



Supervised learning example

Housing price prediction.



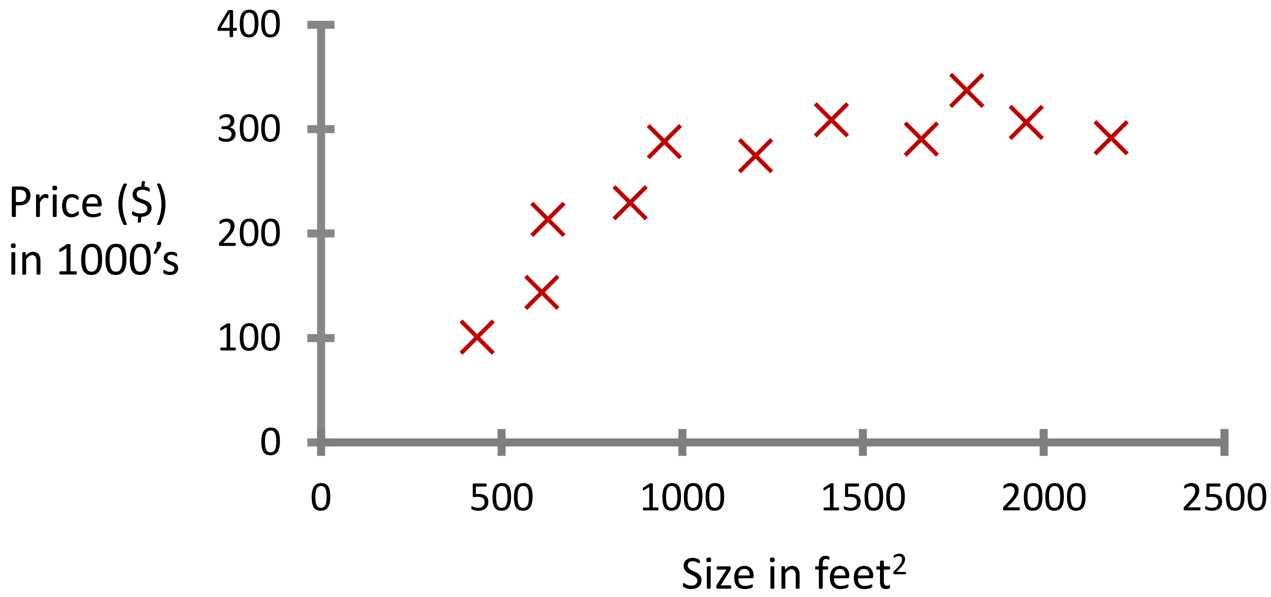
<u>Supervised Learning</u>

"right answers" given

<u>Regression:</u> Predict continuous valued output (price)

Supervised learning example

Housing price prediction.



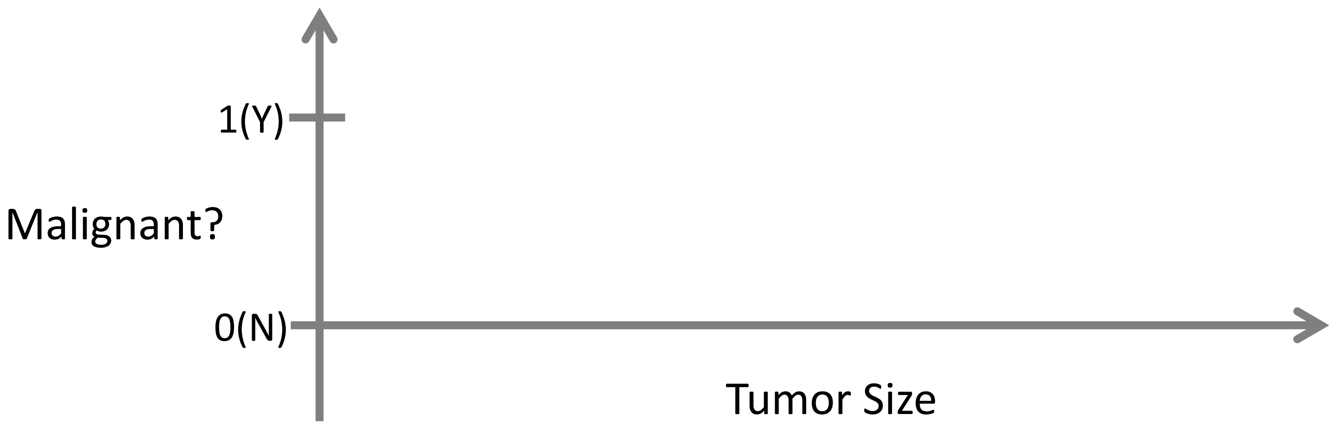
<u>Supervised Learning</u>

"right answers" given

<u>Regression:</u> Predict continuous valued output (price)

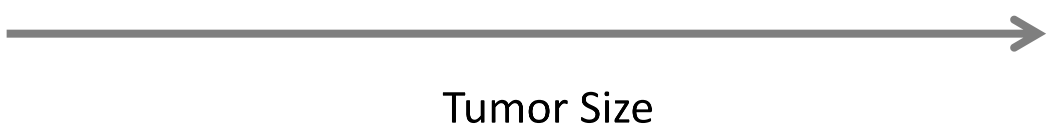
Supervised learning example

Breast cancer (malignant, benign)



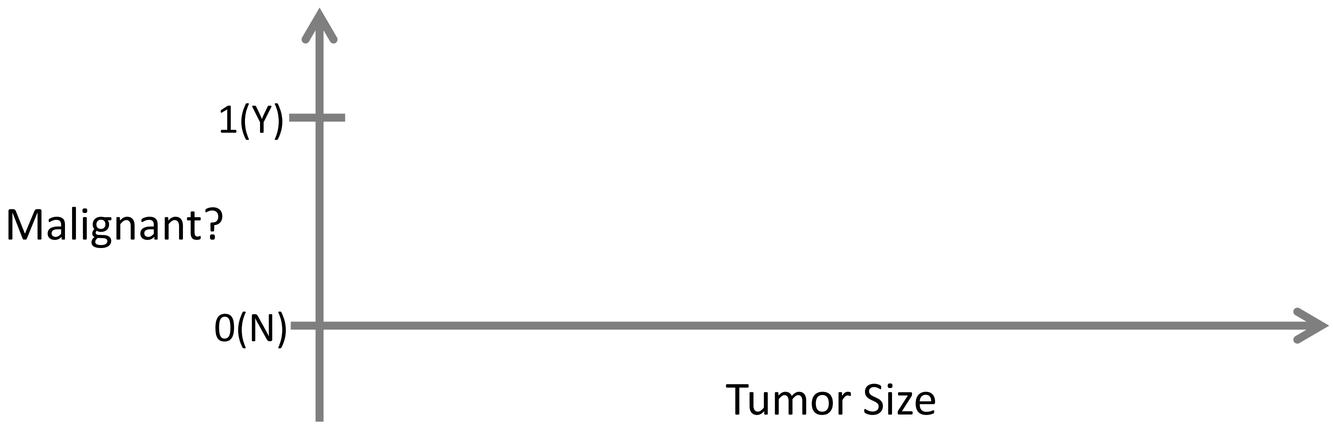
<u>Classification</u>

Discrete valued output (0 or 1)



Supervised learning example

Breast cancer (malignant, benign)



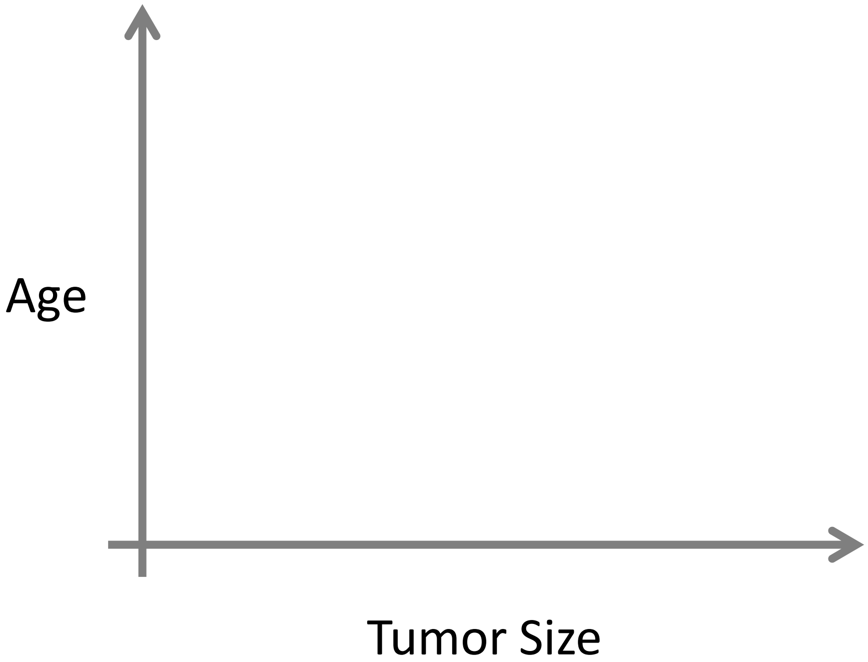
<u>Classification</u>

Discrete valued output (0 or 1)

Only one feature has been used in this case.

Supervised learning example

Breast cancer (malignant, benign): Number of features are more than one

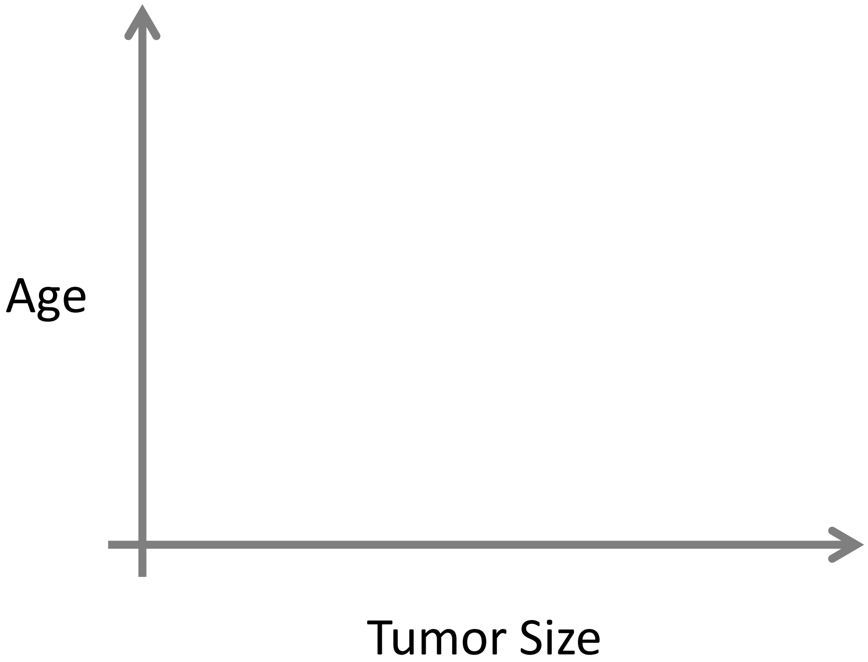


- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape

...

Supervised learning example

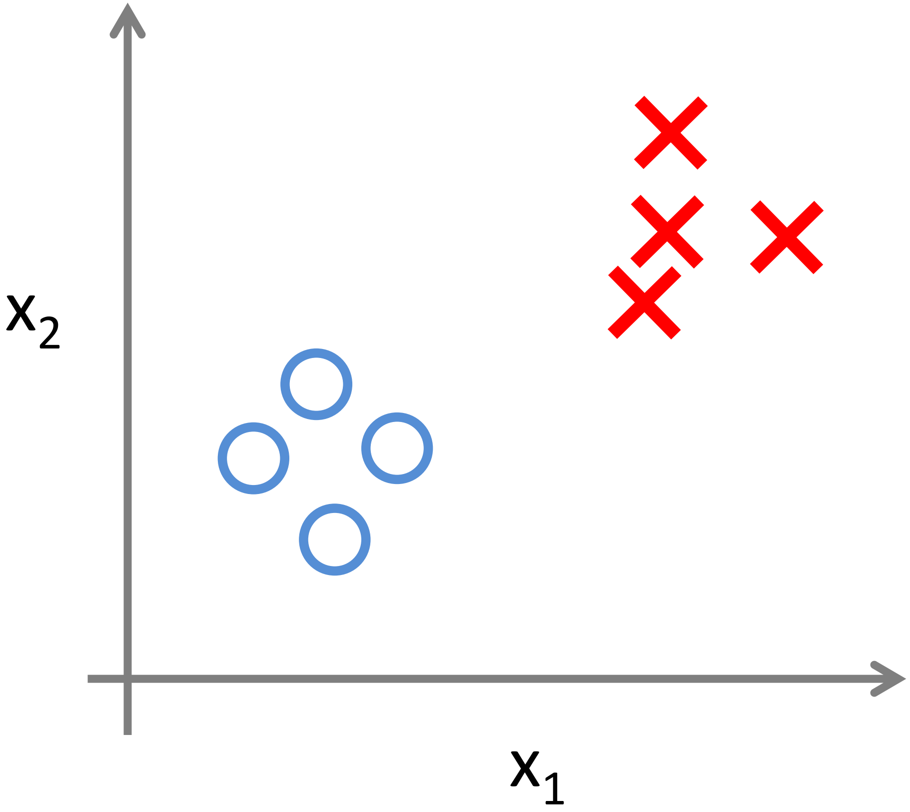
Breast cancer (malignant, benign): Number of features are more than one



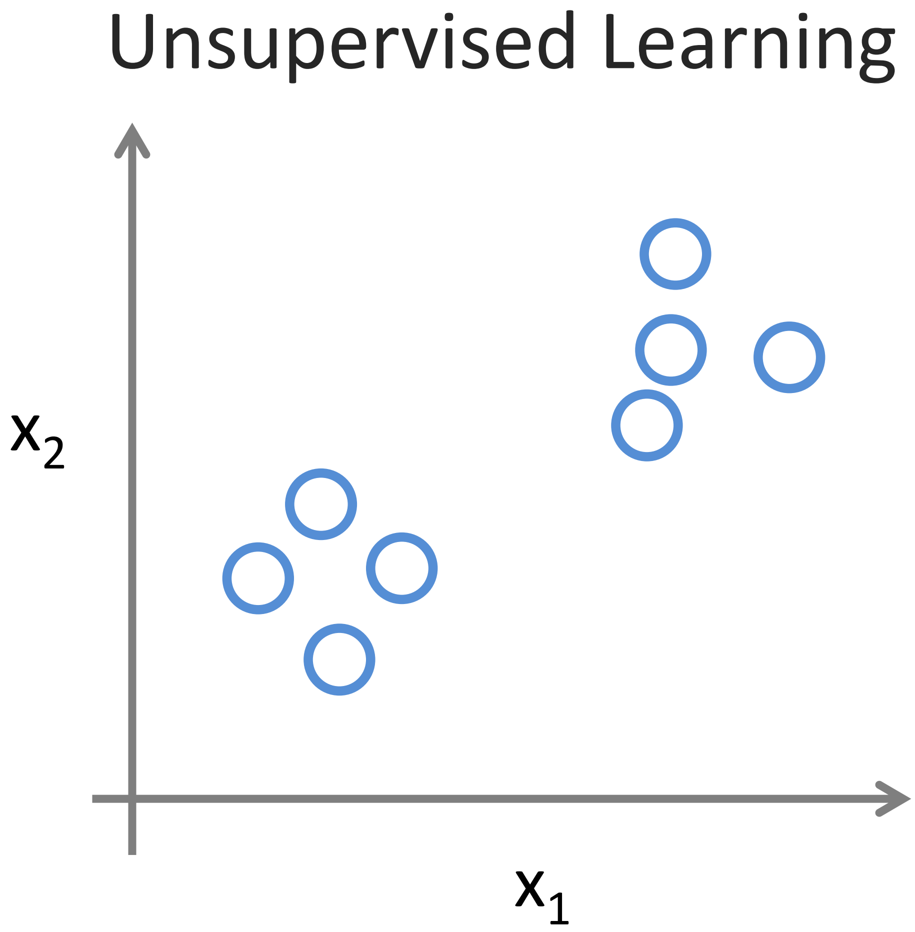
- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape

…

Supervised Learning

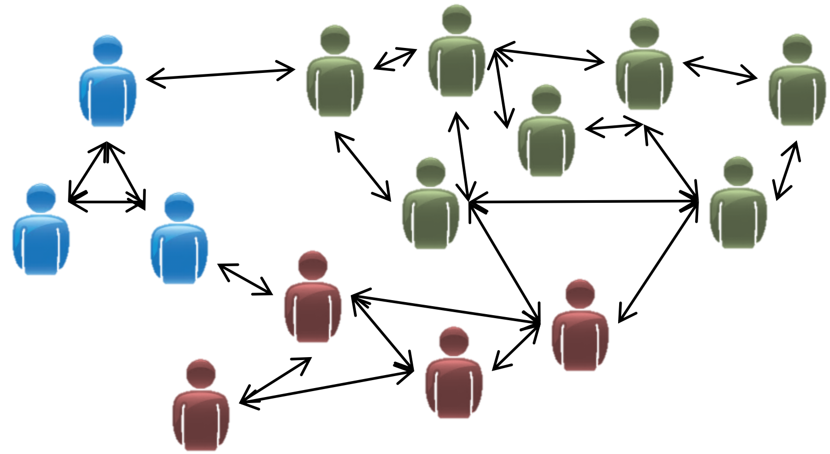


Unsupervised learning example

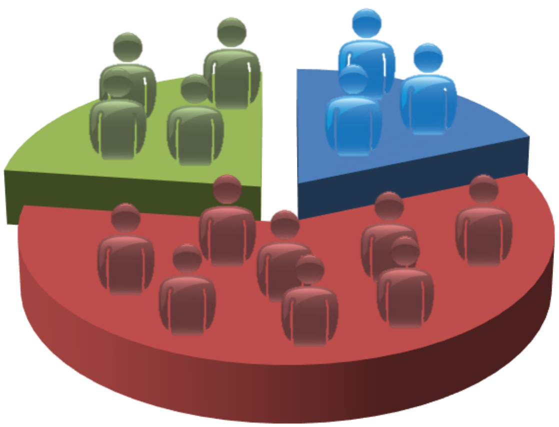


Unsupervised learning example

Organize computing clusters



Social network analysis



Market segmentation

Astronomical data analysis

Questions

Of the following examples, which would you address using an unsupervised learning algorithm?

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

Some examples of tasks best solved by machine learning

- Database mining
  - Large datasets from growth of automation/web.
  - E.g., Web click data, medical records, biology, engineering
- Applications can't program by hand.
  - E.g., Autonomous helicopter, handwriting recognition, most of Natural Language Processing (NLP), Computer Vision.
- Self-customizing programs
  - E.g., Amazon, Netflix product recommendations
- Understanding human learning (brain, real AI).

Some examples of tasks best solved by machine learning

- Recognizing patterns:
  - Object in real scenes
  - Facial identities or facial expressions
  - Handwritten documents
  - Spoken words
- Recognizing anomalies:
  - Unusual sequences of credit card transactions
  - Unusual patterns of sensor readings in a nuclear power plant
- Prediction:
  - Future stock prices or currency exchange rates
  - Which movies will a person like?
- Clustering

Image Classification

Tiger

Giraffe

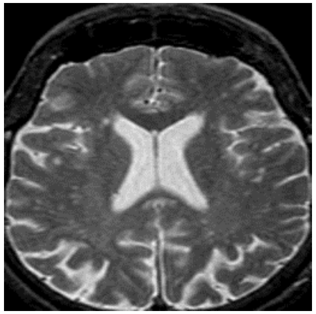
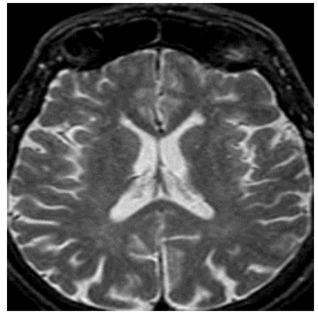
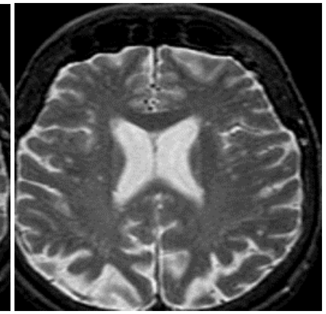
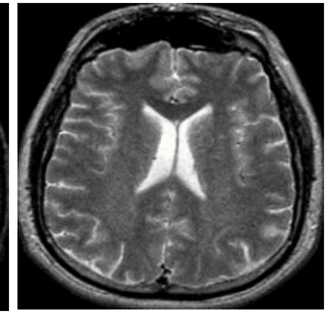
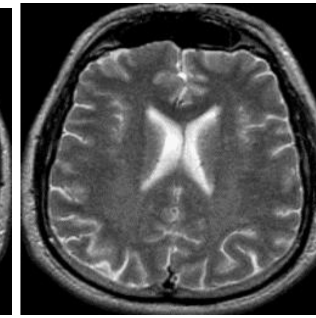
Horse

Bear

Scene Image Classification

Medical Image Classification

Normal brain MR image

Abnormal brain MR image

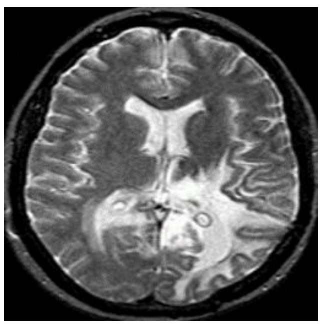
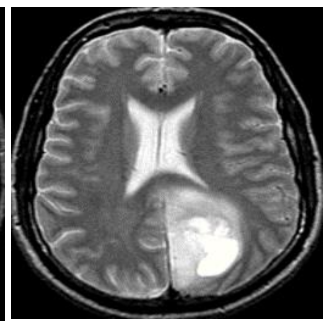
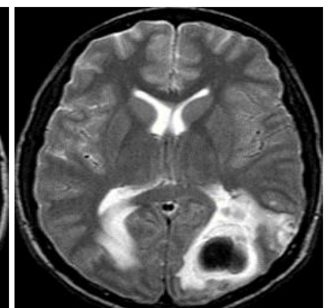
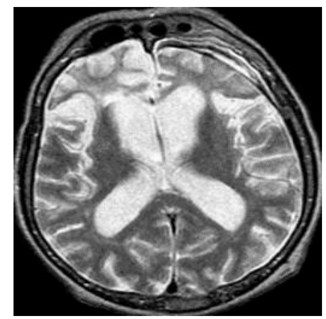
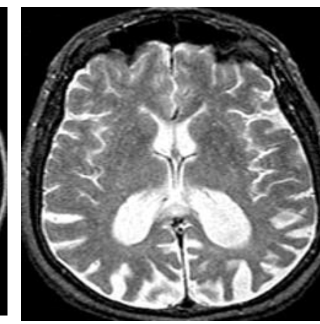






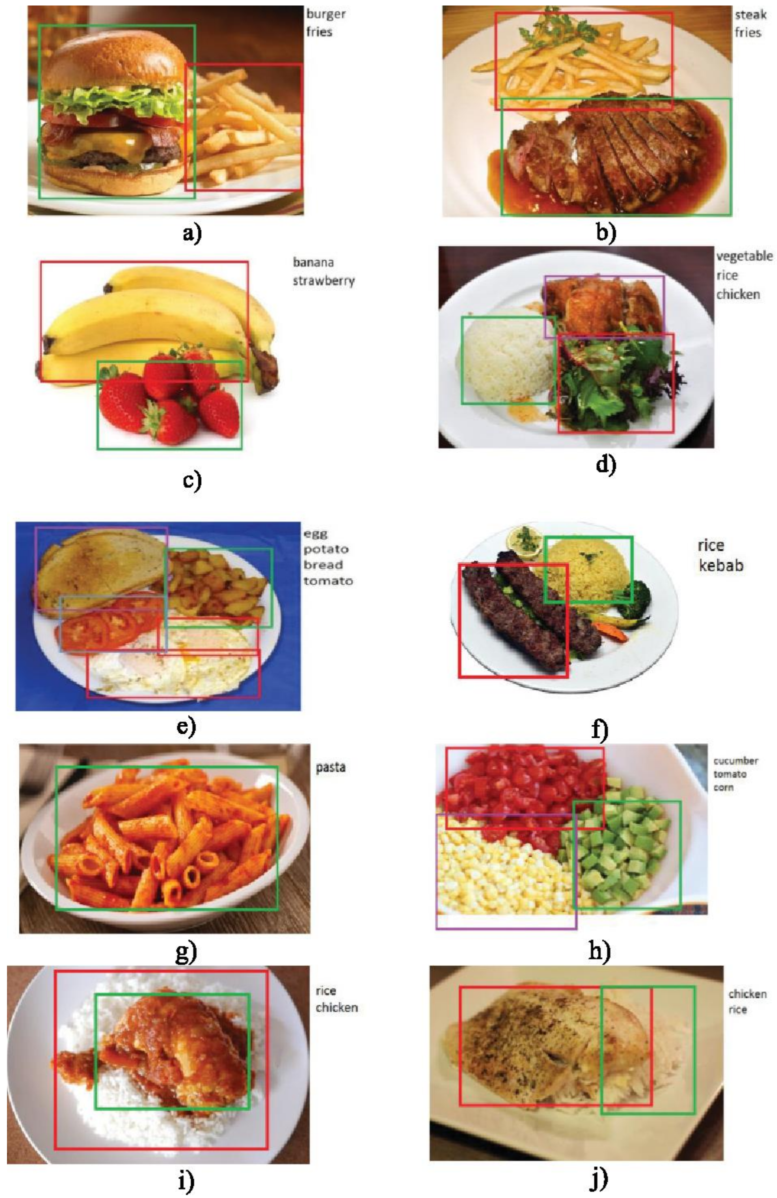
Image Captioning

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

Multi-Food Recognition



Pouladzadeh & Shirmohammadi,
TOMM, 2017

Feature Set

- In a classification task, we are given a pattern and the task is to classify it into one out of *c* classes.

- Supervised (The number of classes *c* is assumed to be known a priori)

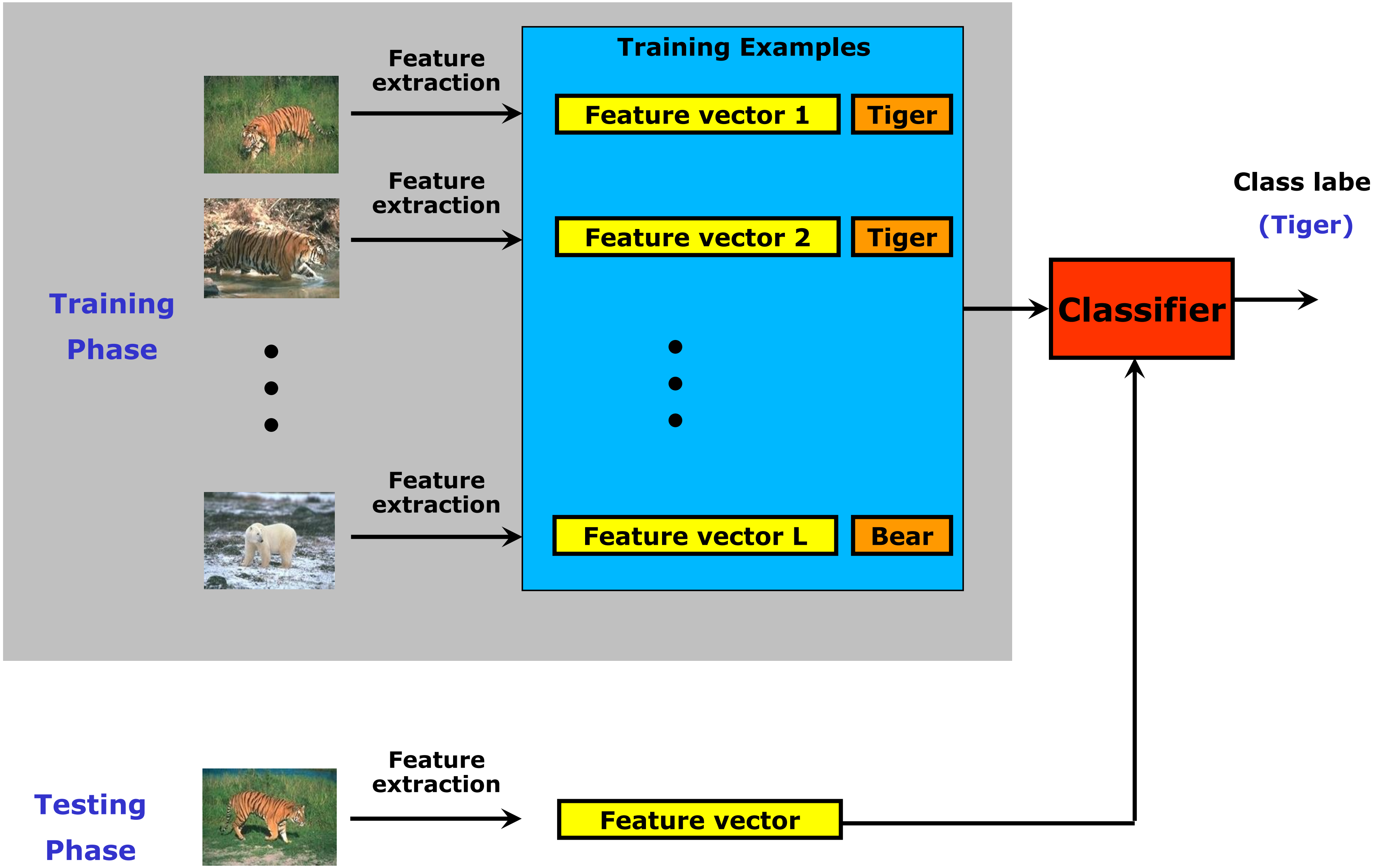
- Each pattern is represented by a vector of features, $x(i), i = 1,2, \ldots, d$ which makes a *d*-dimensional feature vector

$$x = [x(1), x(2), \ldots, x(d)]^T \in \mathbb{R}^d$$

- We assume that each pattern is represented uniquely by a single feature vector and that it can belong to only one class.

Pattern Classification: Example

Static pattern: An example is represented by a vector of features



Evaluation Measures

- Success: actual output = target
- Error: actual output ≠ target
- % error= #errors/#samples

Evaluation Measures

- Training: In training the model is built
- Testing: In testing, the model is applied to the new data.

The goal in building a machine learning algorithm is to perform well on both training and testing data.

- Error on training data is called training error.
- Error on test data is called test error.
- Test error indicates how well model will perform on new data. (**Generalization**)

Performs well on new data ⇒ Good generalization

Test error = Generalization error

Overfitting and Underfitting

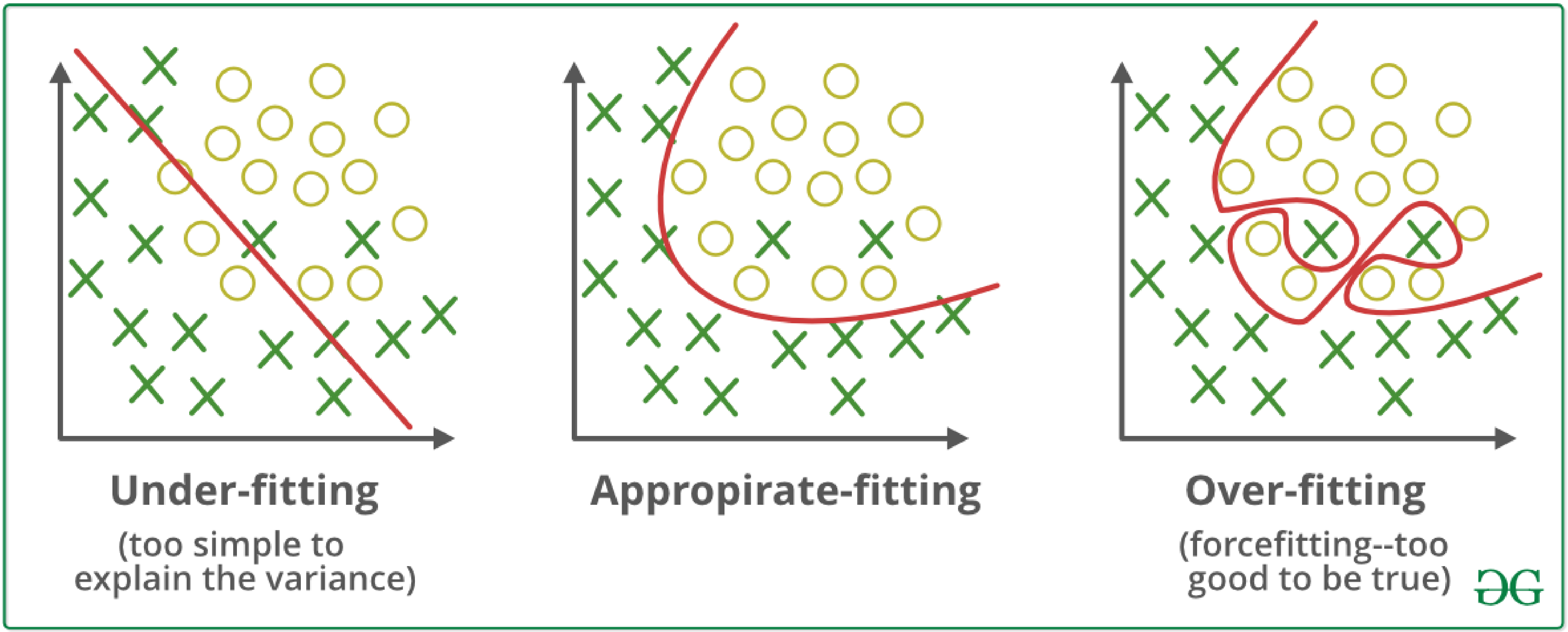
- **Overfitting**: If a model has **low training error** and **high test/generalization error**, then it is called overfitting
- Fit to the noise in the training data

Overfitting = Poor generalization

- **Underfitting**: **high training error** and **high test/generalization error**

- **The cause of the poor performance of a model in machine learning is either overfitting or underfitting the data.**

Overfitting and Underfitting

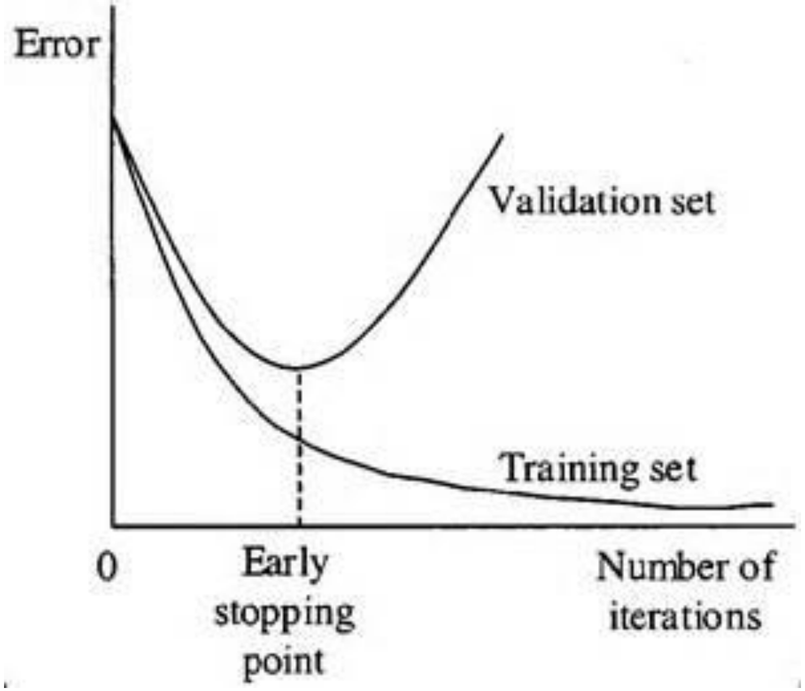


What causes overfitting?

It occurs when a model is too complex i.e., it has too many parameters related to the number of training samples.

So, to avoid overfitting the model need to be kept as simple as possible.

How to prevent Overfitting?



Confusion matrix

Confusion Matrix is a tool to determine the performance of classifier. It contains information about actual and predicted classifications.

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

True Positive (TP): # positive samples correctly identified as positive.
False Negative (FN): #positive samples incorrectly identified as negative
False Positive (FP): # negative samples incorrectly identified as positive
True Negative (TN): # negative samples correctly identified as negative

Accuracy, sensitivity, specificity

sensitivity = true positive rate
= recall
= 1 – false negative rate
= TP / (TP + FN)

specificity = true negative rate
= 1 – false positive rate
= TN / (TN + FP)

accuracy = (TP + TN) / (TP + TN + FP + FN)

precision = TP / (TP + FP)

$$\text{F1 Score} = 2 \times \frac{\textit{Precision} \times \textit{Recall}}{\textit{Precision} + \textit{Recall}}$$

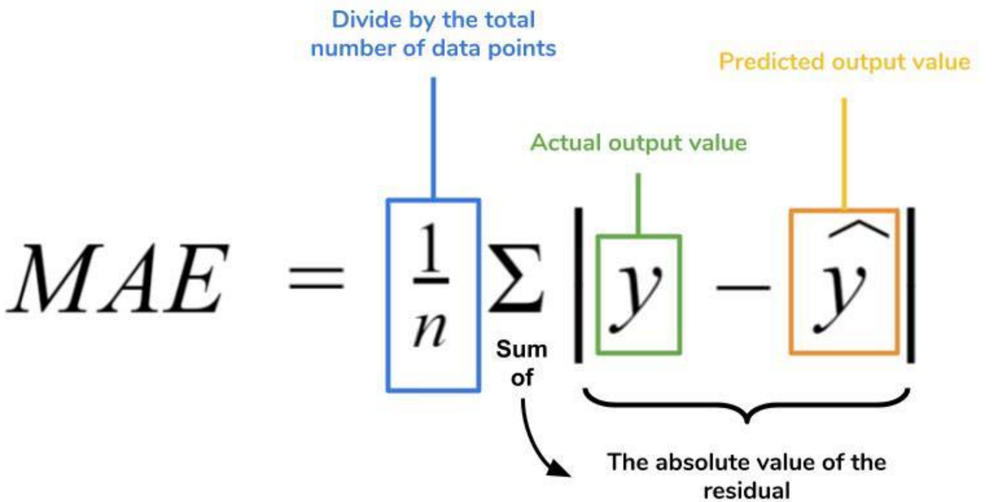
Confusion matrix

		Predicted Class		
		Positive	**Negative**	
Actual Class	**Positive**	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\frac{TP}{(TP+FN)}$
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\frac{TN}{(TN+FP)}$
		Precision $\frac{TP}{(TP+FP)}$	**Negative Predictive Value** $\frac{TN}{(TN+FN)}$	**Accuracy** $\frac{TP+TN}{(TP+TN+FP+FN)}$

https://manisha-sirsat.blogspot.com/2019/04/confusion-matrix.html

Regression Error Metrics

Mean Absolute Error: average of the difference between the Original Values and the Predicted Values. It gives us the measure of how far the predictions were from the actual output.



$$\mathrm{MAE} = \frac{1}{n} \sum_{j=1}^{n} |y_j - \hat{y}_j|$$

https://www.dataquest.io/blog/understanding-regression-error-metrics/

Regression Error Metrics

Mean Square Error (MSE): average of the **square** of the difference between the original values and the predicted values. The advantage of MSE being that it is easier to compute the gradient, whereas Mean Absolute Error requires complicated linear programming tools to compute the gradient

$$MSE = \frac{1}{n} \Sigma \underbrace{\left(y - \widehat{y} \right)^2}_{\text{The square of the difference between actual and predicted}}$$

Root Mean Square Error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Regression Error Metrics

Mean absolute percentage error (MAPE):

Multiplying by 100% converts to percentage

$$MAPE = \frac{\boxed{100\%}}{n} \sum \left| \frac{\overbrace{y - \hat{y}}^{\text{The residual}}}{\boxed{y}} \right|$$

Each residual is scaled against the actual value

Mean percentage error:

$$MPE = \frac{100\%}{n} \sum \left(\frac{y - \hat{y}}{y} \right)$$

Reference Books

- Richard O. Duda, Peter E. Hart, David G. Stork, "Pattern Classification", 2/E, Wiley - Interscience, 2000.
- Christopher M. Bishop :, "Pattern Recogni tion And Machine Learning (Information Science and Statistics)" ,1/E, Springer, January 2008
- T. Hastie , R. Tibshirani, J. H. Friedman:, "The Elements of Statistical Learning",1/E ,Springer, Reprint 3/E, 2003
- Christopher M. Bishop ; "Pattern Recognition and Machine Learning", Springer, 2006
- Shigeo Abe, "Advances in Pattern Recognition", Springer, 2005

Prerequisites

- Linear algebra
- Probability theory
- Statistics
- Programming (MATLAB/Python)

Thank You