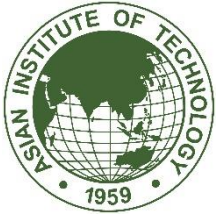


Machine Learning Performance

Dr. Mongkol Ekpanyapong

Machine Learning Paradigm

- Training / Learning
- Testing / Inference

Machine Learning

- Regression
- Classification

Regression

- Regression is a study how to best fit a curve to summarize your data

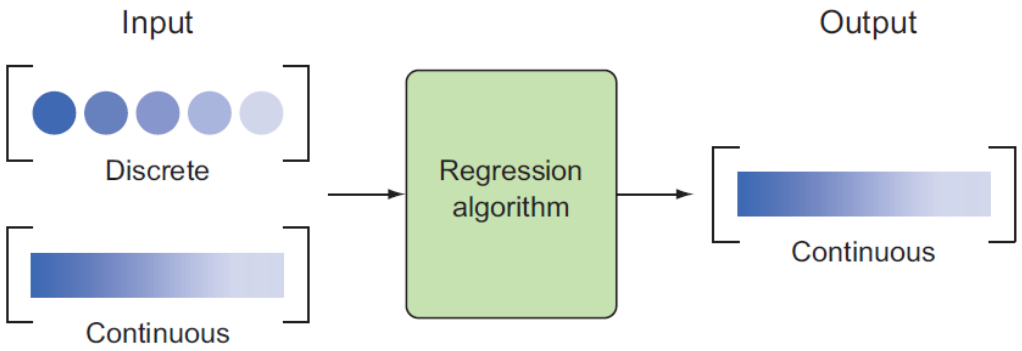
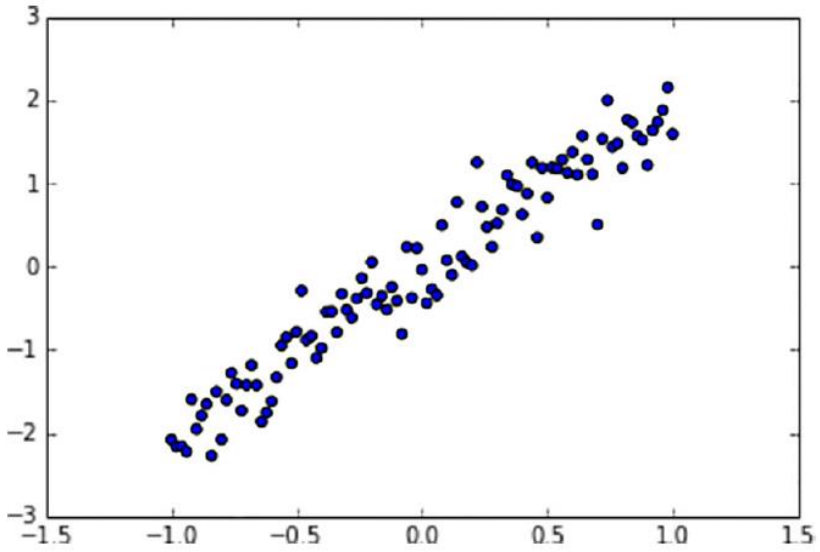
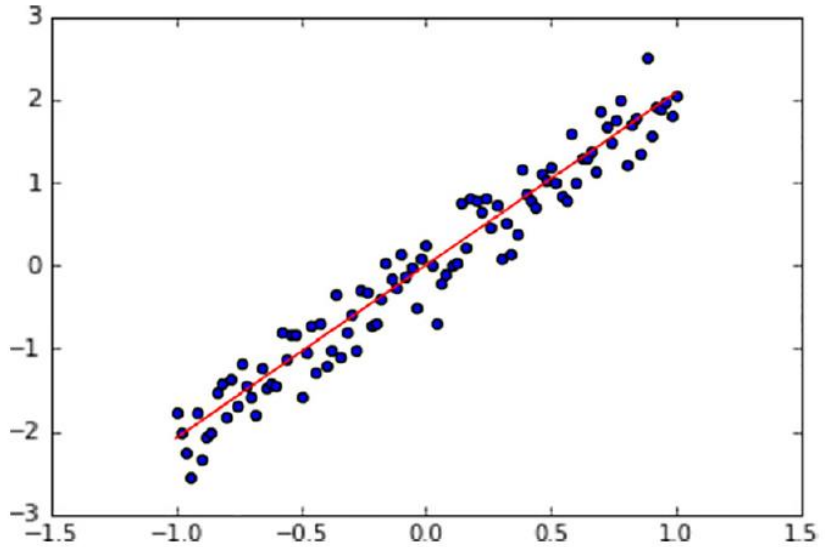


Image from Machine Learning with TensorFlow book

Data Modelling

- Given the data, what is the best function to fit the model:





Classification

- Classification is to assign discrete labels to its inputs
- The input is typically a feature vector
- The output is a class
  - A binary classifier if there are only two class labels
  - A multiclass classifier if there are more than two.

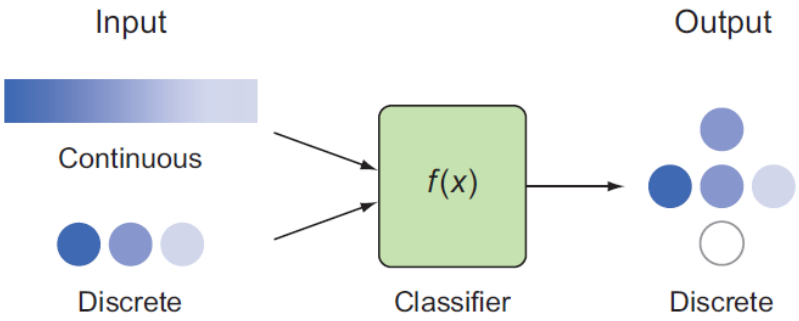


Figure 4.1 A classifier produces discrete outputs but may take either continuous or discrete inputs.

Image from Machine Learning with TensorFlow book

Measuring Performance Terminology

- True Positive(TP): Data items that are correctly predicted
- True Negative(TN): Data items that are correctly predicted as a negative label
- False Positive(FP): Data items that are incorrectly predicted as a positive label
- False Negative(FN): Data items that are predicted as a negative label even though it is positive

Measuring Performance

- Accuracy

$$accuracy = \frac{\#correct}{\#total}$$

- Precision

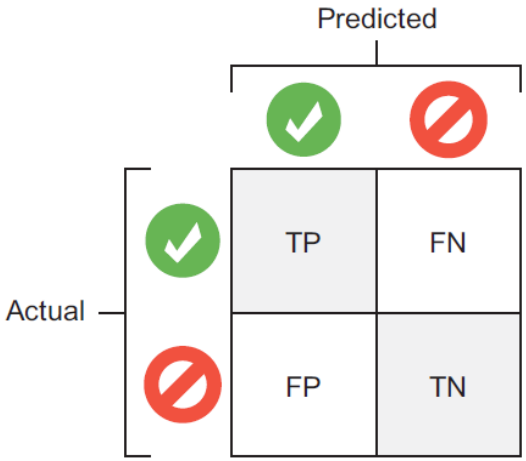
$$precision = \frac{TP}{TP + FP}$$

- Recall

$$recall = \frac{TP}{TP + FN}$$

Confusion Matrix

- Confusion Matrix is a detailed report of machine learning performance



Example for Cat Prediction

Confusion matrix		Predicted	
		Cat	Dog
Actual	Cat	30	20
	Dog	10	40

- Accuracy = 70/100
- Precision = 30/40
- Recall = 30/50

If you have to make 100% recall for cat prediction, what will you do?

100% recall => don't miss any cat image

Always predict everything as a cat

=> it will create a lot of false positive

Other terms

- Fault Acceptance Rate (FAR)

  FAR = # of false claims/ # attempts

- Fault Rejection Rate (FRR)

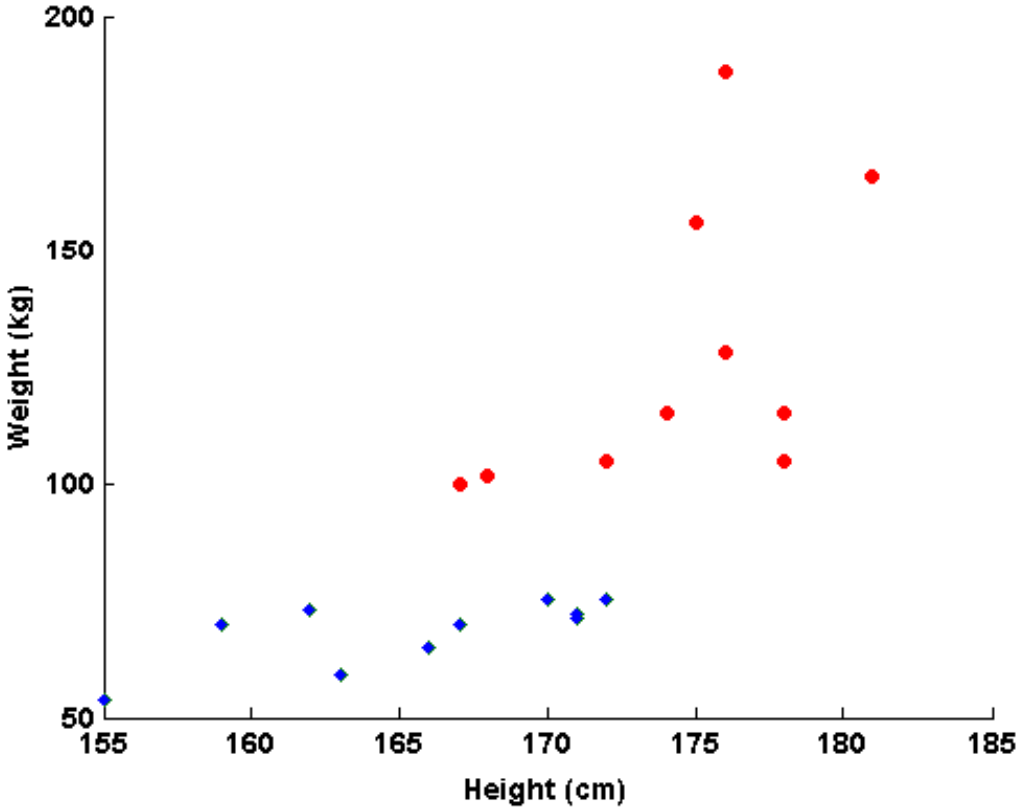
  FRR = # of true claims rejected / #attempts

Patterns and pattern classes

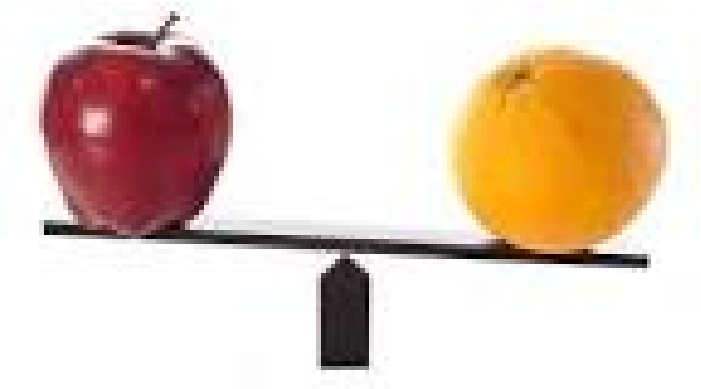
- A *pattern* can be defined as an arrangement of *descriptors* or *features*.
  - Patterns are usually encoded in the form of feature vectors, strings, or trees.
- A *class* is a set of patterns that share some common properties.
  - An ideal class is one in which its members are very similar to one another (i.e., the class has **high intra-class similarity**) and yet significantly different from members of other classes (i.e., **inter-class differences are significant**)

Patterns and pattern classes

- Sumo wrestlers and table tennis players



Apple and Orange Comparison



Patterns and pattern classes

- Data preprocessing
  - **Noise removal**: data samples that deviate too far from the average value for a class are removed, under the rationale that: (a) there may have been a mistake while measuring (or extracting) that particular sample; (b) the sample is a poor example of the underlying structure of the class.
  - **Normalization**: feature vectors may need to be normalized before distance, similarity, and probability calculations take place.
  - **Insertion of missing data**: (optional).

Training and test sets

- The process of development and testing of pattern classification algorithms usually requires that the dataset be divided in two subgroups:
  - **training set**: used for algorithm development and fine-tuning
  - **test set**: used to evaluate the algorithm's performance.
- The test set contains a small (typically 20% or less), representative subsample of the dataset, selected manually or automatically.
- The size of the training set and the method used to build it are often dependent on the selected pattern classification technique.
- The goal of having two separate sets is to avoid bias in reporting the success rates of the approach.

Confusion matrix

- A 2D array of size $K \times K$ (where K is the total number of classes) used to report raw results of classification experiments.
- The value in row i, column j indicates the number of times an object whose true class is i was labeled as belonging to class j.
- The main diagonal of the confusion matrix indicates the number of cases where the classifier was successful; a perfect classifier would show all off-diagonal elements equal to zero.

Confusion matrix

- Example :

Predict

True class

	ω_1	ω_2	ω_3	ω_4
ω_1	97	0	2	1
ω_2	0	89	10	1
ω_3	0	0	100	0
ω_4	0	3	5	92

- Example :
  - Overall error rate: 5.5 % (success rate = 94.5%)

Precision and recall

- Certain image processing applications, notably image retrieval, have as their goal to retrieve *relevant* images while not retrieving *irrelevant* ones.
- The measures of performance used in image retrieval borrow from the field of (document) information retrieval and are based on two primary figures of merit: *precision* and *recall*.
  - **Precision** is the number of relevant documents retrieved by the system divided by the total number of documents retrieved (i.e., true positives plus false alarms).
  - **Recall** is the number of relevant documents retrieved by the system divided by the total number of relevant documents in the database (which should, therefore, have been retrieved).

Precision and recall

- Example :

	ω_1	ω_2	ω_3	ω_4
ω_1	97	0	2	1
ω_2	0	89	10	1
ω_3	0	0	100	0
ω_4	0	3	5	92

$$P = \frac{tp}{tp + fp}$$

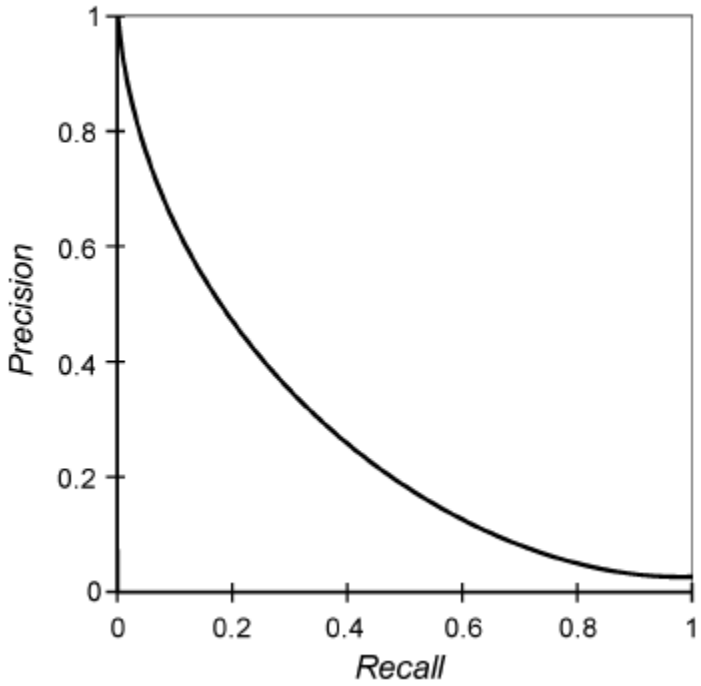
$$R = \frac{tp}{tp + fn}$$

$$P_1 = 97/(97 + 0 + 0 + 0) = 100\%$$
$$P_2 = 89/(0 + 89 + 0 + 3) = 96.74\%$$
$$P_3 = 100/(2 + 10 + 100 + 5) = 85.47\%$$
$$P_4 = 92/(1 + 1 + 0 + 92) = 97.87\%$$
$$R_1 = 97/(97 + 0 + 2 + 1) = 97\%$$
$$R_2 = 89/(0 + 89 + 10 + 1) = 89\%$$
$$R_3 = 100/(0 + 0 + 100 + 0) = 100\%$$
$$R_4 = 92/(0 + 3 + 5 + 92) = 92\%$$

Precision and recall

- Precision can be interpreted as a measure of *exactness*, whereas recall provides a measure of *completeness*.
  - A perfect precision score of 1.0 means that every retrieved document (or image, in our case) was relevant, but does not provide any insight as to whether all relevant documents were retrieved.
  - A perfect recall score of 1.0 means that all relevant images were retrieved, but says nothing about how many irrelevant images might have also been retrieved.

Precision and recall graph



- P-R graph
  - Obtained by calculating the precision at various recall levels.
  - The ideal P-R graph shows perfect precision values at every recall level until the point where all relevant documents (and only those) have been retrieved; from that point on it falls monotonically until the point where recall reaches one.

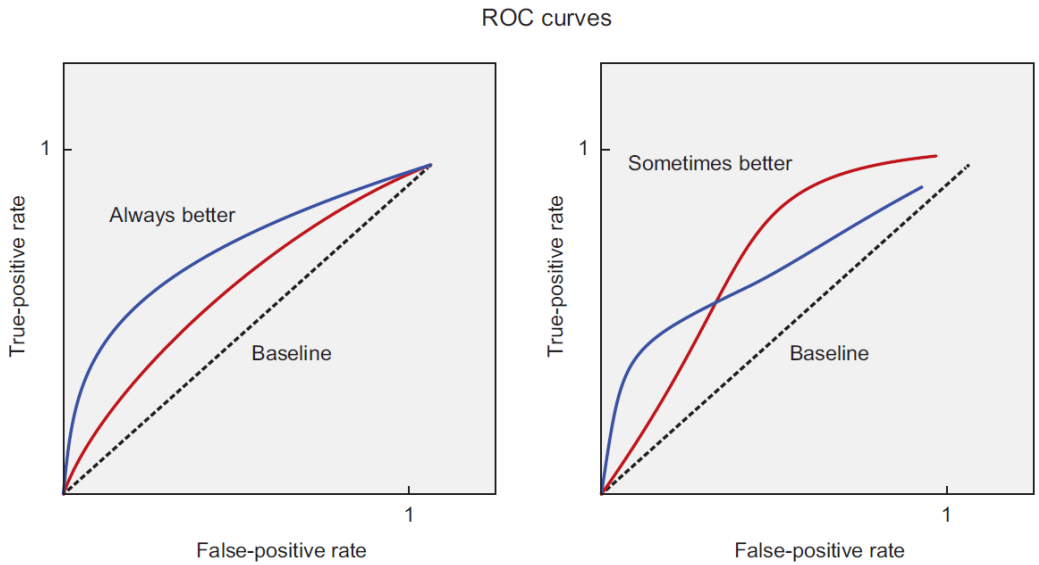
Precision and recall

- F1: a more compact representation of the precision and recall properties of a system.

$$F1 = 2 \times \frac{precision \times recall}{precision + recall}$$

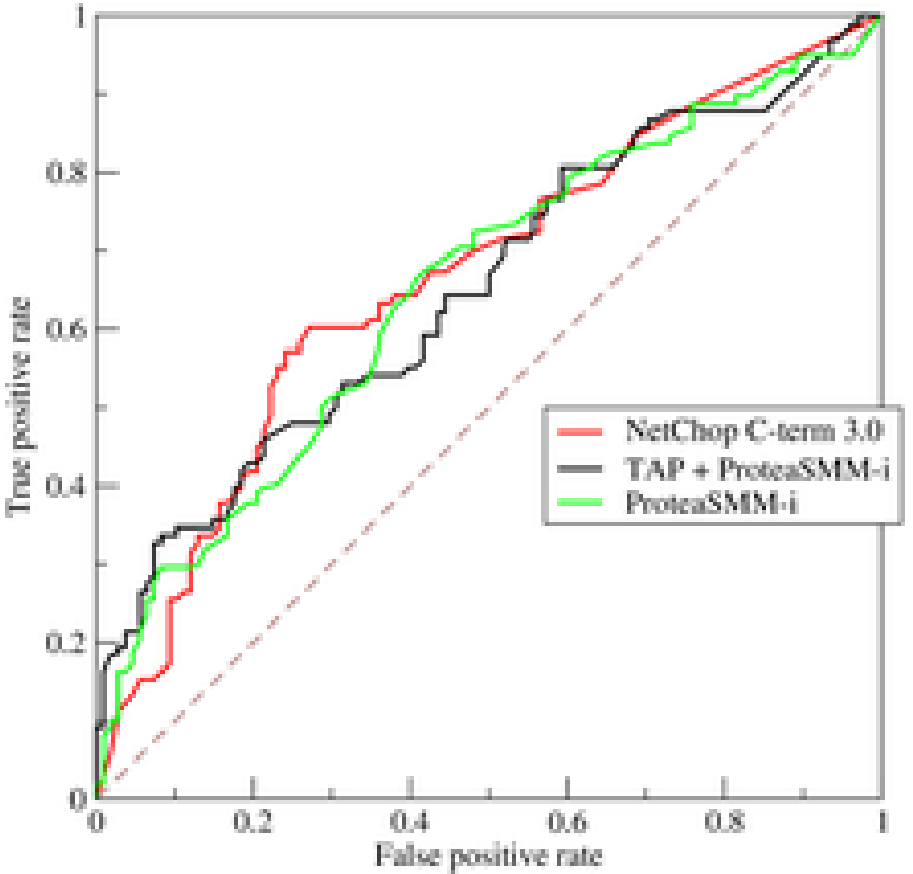
Receiver Operating Characteristic curve (ROC)

- The ROC curve is a plot that lets you compare the trade-offs between false positives and true positives
- Example of two machine learning models's performance comparison



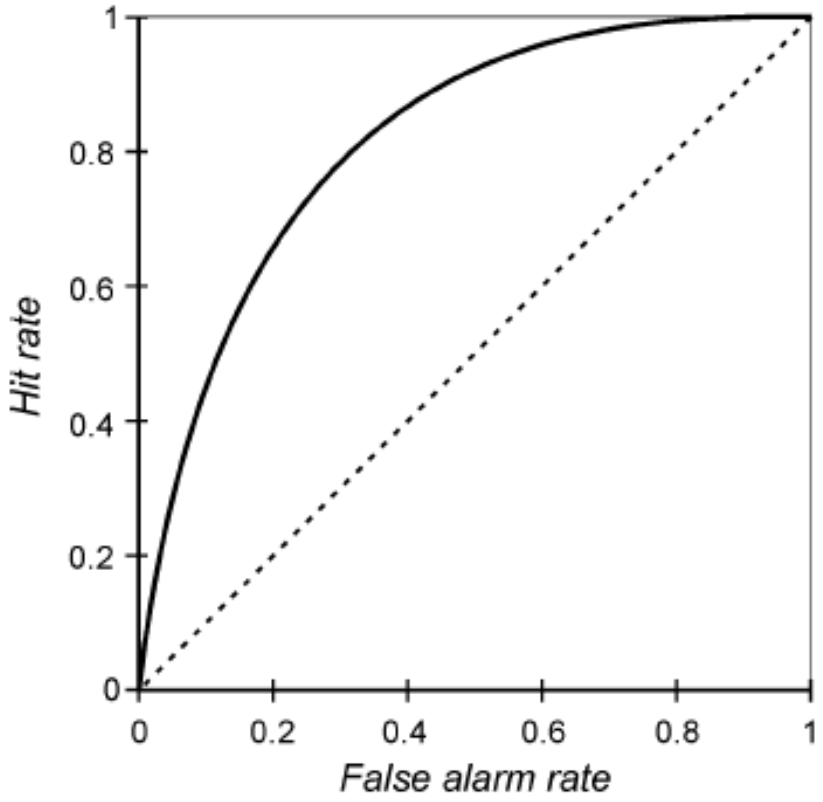
Receiver Operating Characteristics (ROC) Curve

- The curve is created by plotting the true positive rate against the false positive rate



Hit rates, false alarm rates, and ROC curves

- Example of ROC curve

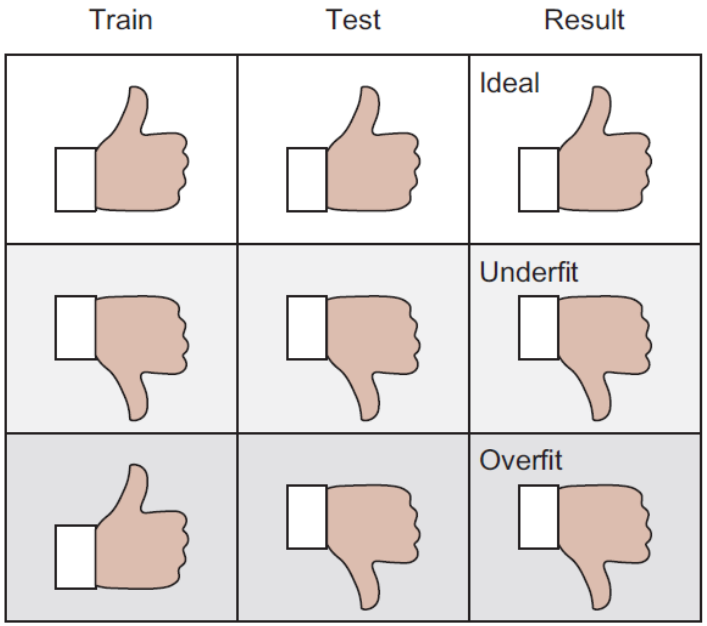


Area-Under-Curve (AUC)

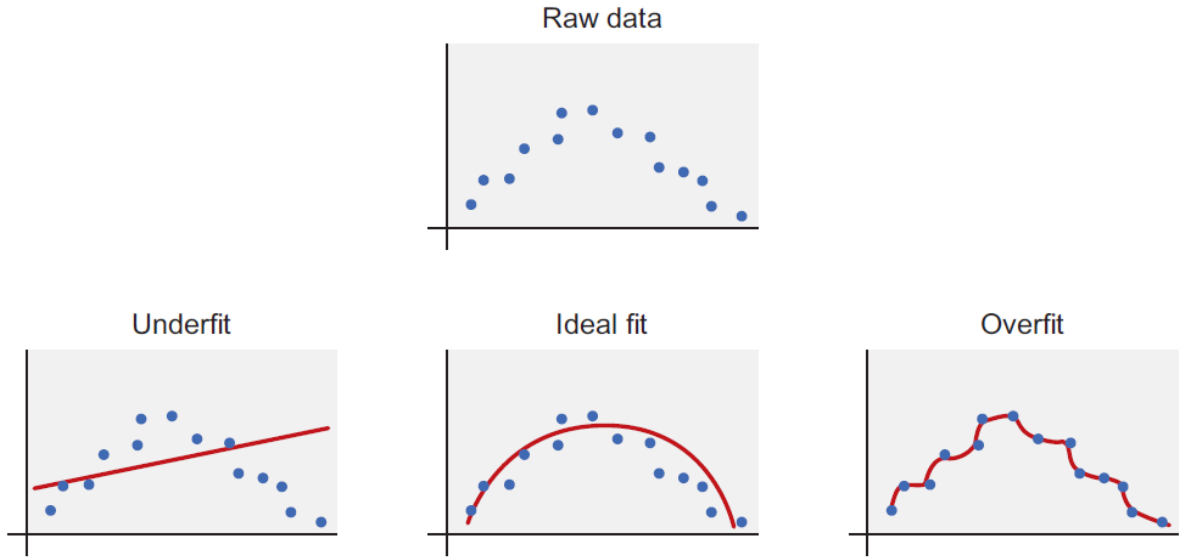
- We can also use AUC as another metric
- With, AUC is higher than 0.9 is a good machine learning model

Overfitting/Underfitting

- With the machine learning, there are three scenarios that can happen



Regression Model



Ranked accuracy

Class Label	Probability
Airplane	0.0%
Automobile	0.0%
Bird	2.1%
Cat	0.03%
Deer	0.01%
Dog	0.56%
Frog	**97.3%**
Horse	0.0%
Ship	0.0%
Truck	0.0%

Class Label	Probability
Airplane	**1.1%**
Automobile	**38.7%**
Bird	0.0%
Cat	0.5%
Deer	0.0%
Dog	0.4%
Frog	0.11%
Horse	**1.4%**
Ship	**2.39%**
Truck	**55.4%**

Rank-5 accuracy

- Step 1: Compute the class label probability for each input image
- Step 2: Sort the predicted class label probabilities in descending order with higher probability are placed at the front of the list
- Step 3: Determine if the ground-truth label exists in the top-5 predicted labels
- Step 4: Tally the number of times where Step 3 is true

Rank-1 and Rank-5 Accuracy

Siberian husky

Eskimo dog

Two different classes in ImageNet from 1,000 classes

Why Rank-5 Accuracy?

- With large classes, it is not easy to get high accuracy of rank-1
- Rank-5 can help see the performance even though we see less improvement in rank-1
- Some class are very similar

Questions?

AI Application

https://quickdraw.withgoogle.com/