Perceptron Networks and Applications

M. Ali Akcayol
Gazi University
Department of Computer Engineering

Content

- Applications
  - Classification
  - Clustering
  - Pattern association
  - Function approximation
  - Forecasting
- Evaluation of neural networks
  - Quality of results
  - Generalizability
  - Computational resources

Applications

- Perception, recognition, memory, conscious thought, dreams, sensorimotor control are desired to be simulated by ANNs.
- The tasks performed using neural networks can be classified as supervised or unsupervised.
- In supervised learning, a teacher is available to indicate whether a system is performing correctly.
- Unsupervised learning, where no teacher is available and learning must rely on guidance obtained heuristically by the system.
- Supervised learning is provided by classification problems, whereas unsupervised learning is provided by clustering problems.

Content

- Applications
  - Classification
  - Clustering
  - Pattern association
  - Function approximation
  - Forecasting
- Evaluation of neural networks
  - Quality of results
  - Generalizability
  - Computational resources

Applications

Classification

- Classification, the assignment of each object to a specific class, has an importance in many areas.
- A training set consisting of sample patterns that are representative of all classes with class membership information for each pattern.
- The rules for membership in each class are deduced and created by a classifier.
- The classifier can then be used to assign new patterns to their respective classes.
- Neural networks have been used to classify samples, map input patterns to different classes.
- The each output node can used for one class.

Applications

Classification

- For two-class problems, feedforward networks with a single output node are adequate.
- An output value close to 1 (>0.9) is considered to indicate one class, while an output value close to 0 (<0.1) indicates the other class.
- Neural networks have been used successfully in a large number of practical classification tasks:
  - Recognizing printed or handwritten characters
  - Classifying loan applications into credit-worthy and non-credit-worthy groups
  - Analyzing sonar and radar data to determine the nature of the source of a signal

Content

- Applications
  - Classification
  - Clustering
  - Pattern association
  - Function approximation
  - Forecasting
- Evaluation of neural networks
  - Quality of results
  - Generalizability
  - Computational resources

Applications

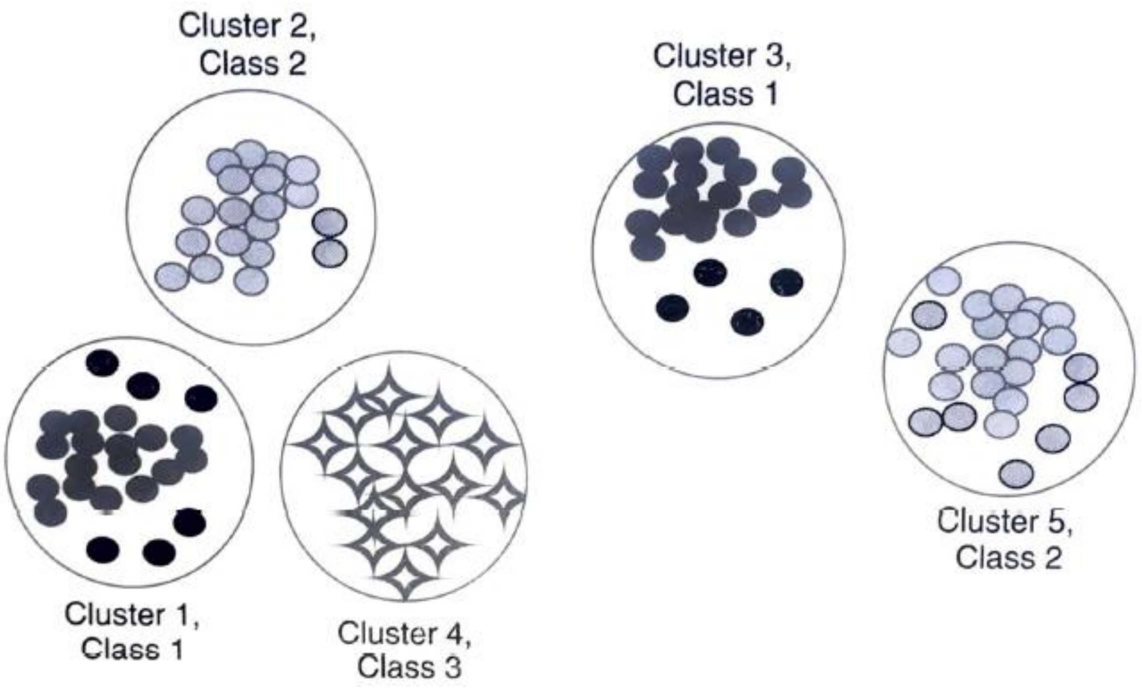
Clustering

- Clustering requires grouping together objects that are similar to each other.
- In clustering problems, all data are created using distance relationships that can be derived from the sample descriptions.
- Most clustering algorithms are based on some distance measure.
- Similar objects have nearly the same values for different features.
- Generally, the main aim is to be minimize intra-cluster distances while maximizing inter-cluster distances.
- Intra-cluster distance can be measured using the distance between different samples from the center (clustroid).
- Inter-cluster distance can be measured using the distance between the centers of different clusters.

Applications

Clustering

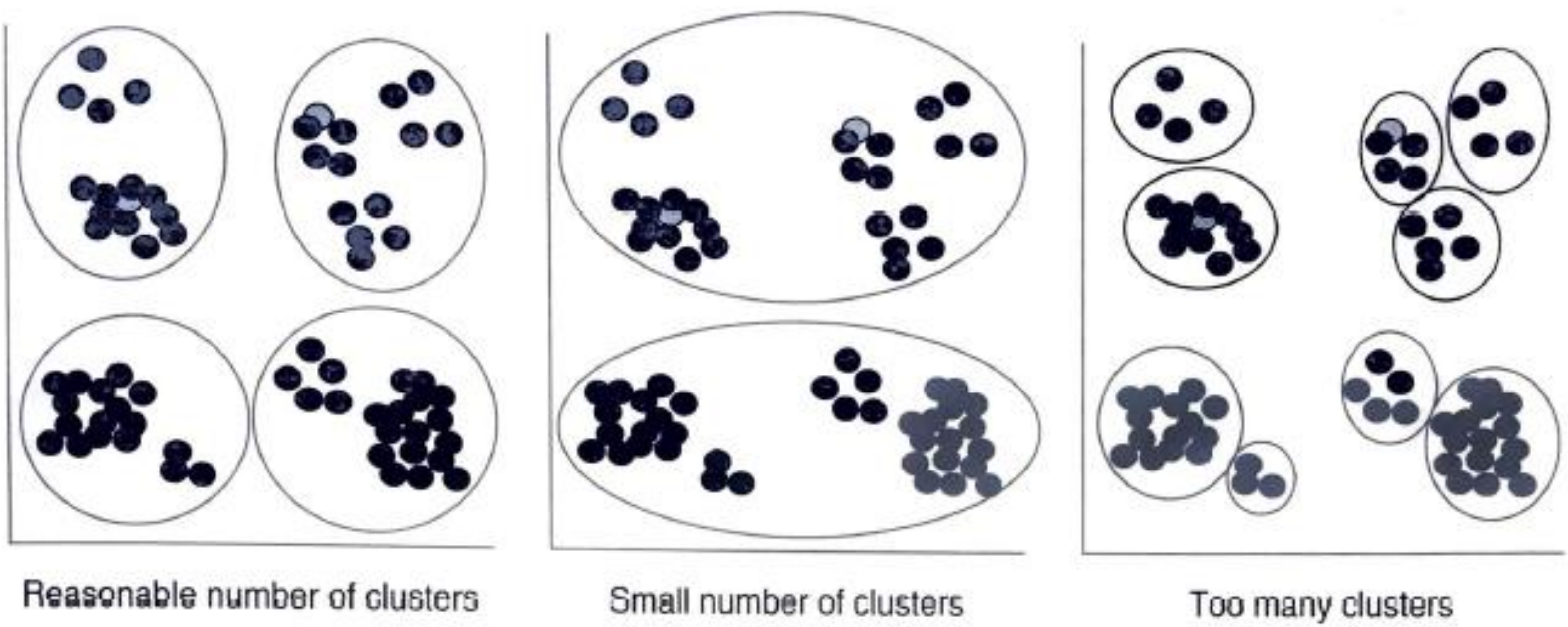
- The number of clusters depends on the problem.



Applications

Clustering

- The first clusters are preferable since it has neither too many nor too few clusters.



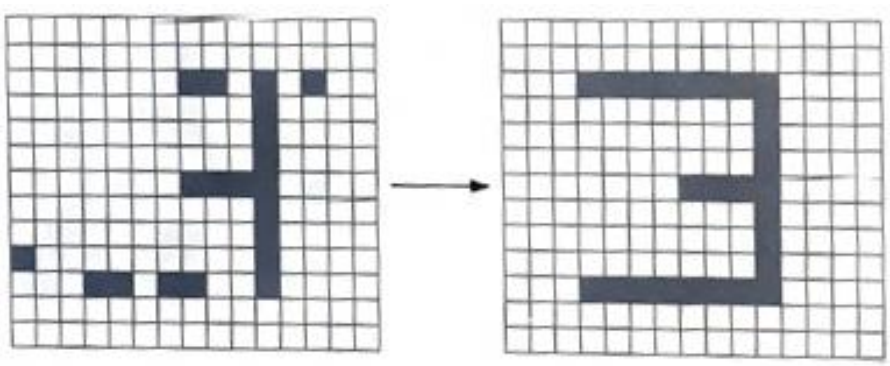
Content

- Applications
  - Classification
  - Clustering
  - Pattern association
  - Function approximation
  - Forecasting
- Evaluation of neural networks
  - Quality of results
  - Generalizability
  - Computational resources

Applications

Pattern association

- Another important task that can be performed by neural networks is pattern association.
- The input is presumed to be a corrupted, noisy, or partial version of the desired output pattern.
- The output pattern may be any arbitrary pattern.
- The left input is corrupted and has missing information, ANN generates the possible correct output.



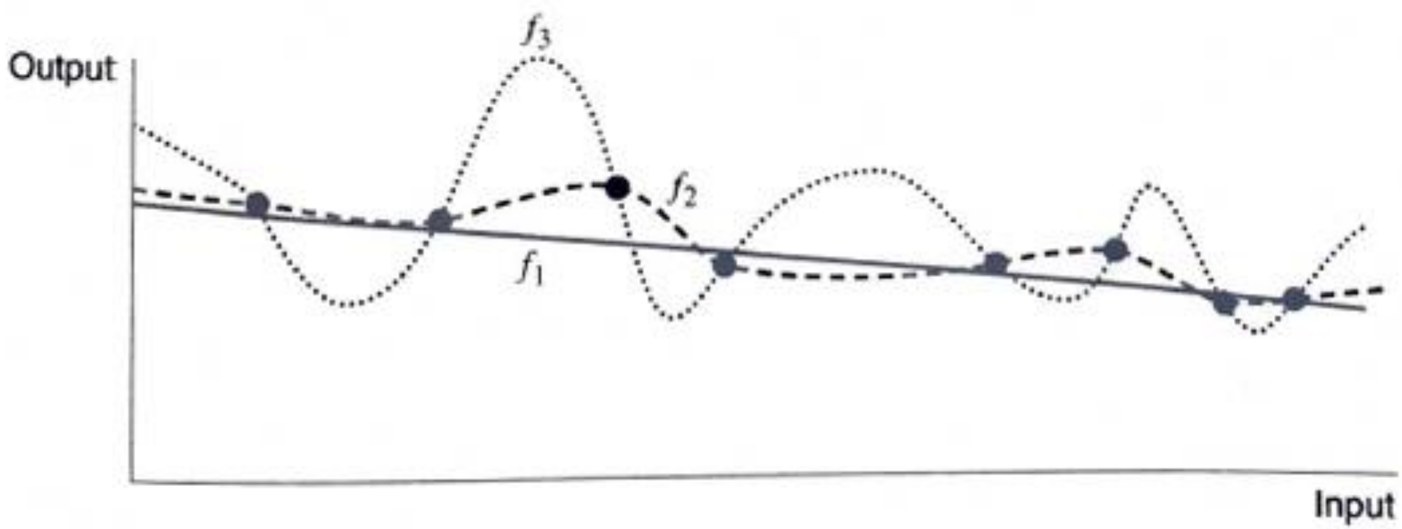
Content

- Applications
  - Classification
  - Clustering
  - Pattern association
  - Function approximation
  - Forecasting
- Evaluation of neural networks
  - Quality of results
  - Generalizability
  - Computational resources

Applications

Function approximation

- Many ANN models can be used as functions mapping some numerical inputs to numerical outputs.
- Function approximation is the task of learning or constructing a function that generates approximately the same outputs from inputs.
- Many different functions can be used to obtain the same finite set of samples.



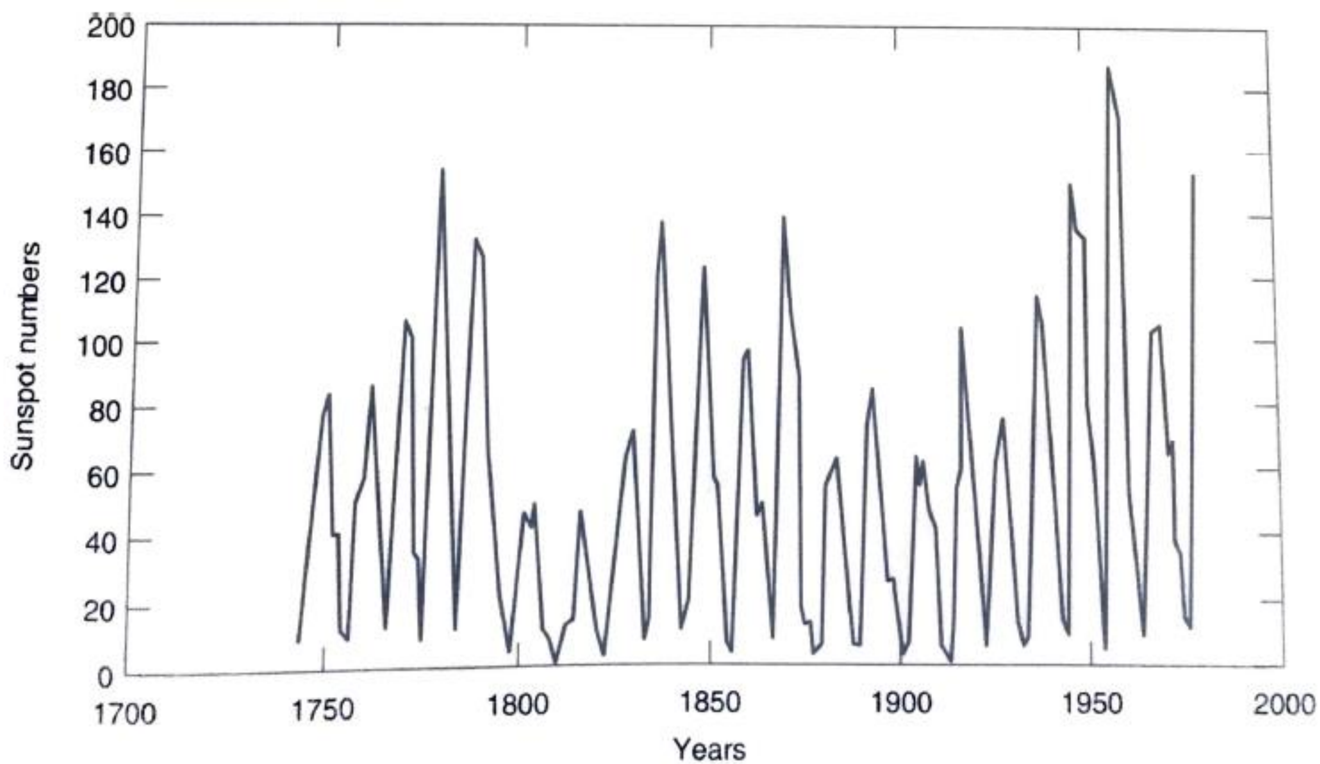
Content

- Applications
  - Classification
  - Clustering
  - Pattern association
  - Function approximation
  - Forecasting
- Evaluation of neural networks
  - Quality of results
  - Generalizability
  - Computational resources

Forecasting

- There are many real-life problems in which future events must be predicted on the basis of past history.
- Perfect prediction is hardly ever possible yet.
- A time series is a sequence of values measured over time.

$$S = \{v(t) : 1 \leq t \leq N\}$$



Applications

Forecasting

- For a network that is to make predictions based upon d most recent values of the variable.
- We extract from S a training set of (d + 1)-tuples.
- Each such tuple contains d + 1 consecutive elements from S.
- The first d components represent inputs and the last one represents the desired output.
- In training phase a d-tuple is presented to the network as input
- The network attempts to predict the next value in the time sequence.
- Feedforward as well as recurrent networks have been used for forecasting.

Content

- Applications
  - Classification
  - Clustering
  - Pattern association
  - Function approximation
  - Forecasting
- Evaluation of neural networks
  - Quality of results
  - Generalizability
  - Computational resources

Evaluation of neural networks

- A human learner's capabilities can be evaluated in the following ways.
  - How well does the learner perform on the data on which the learner has been trained,
  - How well does the learner perform on new data not used for training the learner?
  - What are the computational resources (time, space, effort) required by the learner?
- The performance of neural networks can also be evaluated using the same criteria.

Content

- Applications
  - Classification
  - Clustering
  - Pattern association
  - Function approximation
  - Forecasting
- Evaluation of neural networks
  - Quality of results
  - Generalizability
  - Computational resources

Evaluation of neural networks

Quality of results

- The performance of a neural network is frequently determined in terms of an error measure.
- The Euclidean distance is most often used for error measure.

$$\left(\sum_i (d_i - o_i)^2\right)^{1/2}$$

where d_i is the ith element of desired output vector, o_i is ith element of output vector.

- Other measures are Manhattan and Hamming distance.

$$\sum_i |d_i - o_i|$$

- These measures are generalized to the Minkowski norms,

$$\left(\sum_i |d_i - o_i|^k\right)^{1/k} \quad k > 0$$

Evaluation of neural networks

Quality of results - classifications

- Different error measures should be used for different kind of problems.
- In classification problems, error measure is the fraction of misclassified samples.

$$E = \frac{\text{Number of misclassified samples}}{\text{Total number of samples}}$$

- Accuracy measures how many predictions are correct over all the predictions.

$$Accuracy = \frac{\text{Number of correct classifications}}{\text{Total number of test cases}}$$

- Accuracy gives a ratio without telling what types of errors.
- Accuracy is significantly affected by imbalanced classes.

Evaluation of neural networks

Quality of results - classifications

- Confusion Matrix is used to determine the performance of a classifier.

		Predicted Class	
		Positive	**Negative**
Actual Class	**Positive**	True Positive (TP)	False Negative (FN) **Type II Error**
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)

True Positive (TP) : Observation is positive, and is predicted to be positive.

False Negative (FN) : Observation is positive, but is predicted negative.

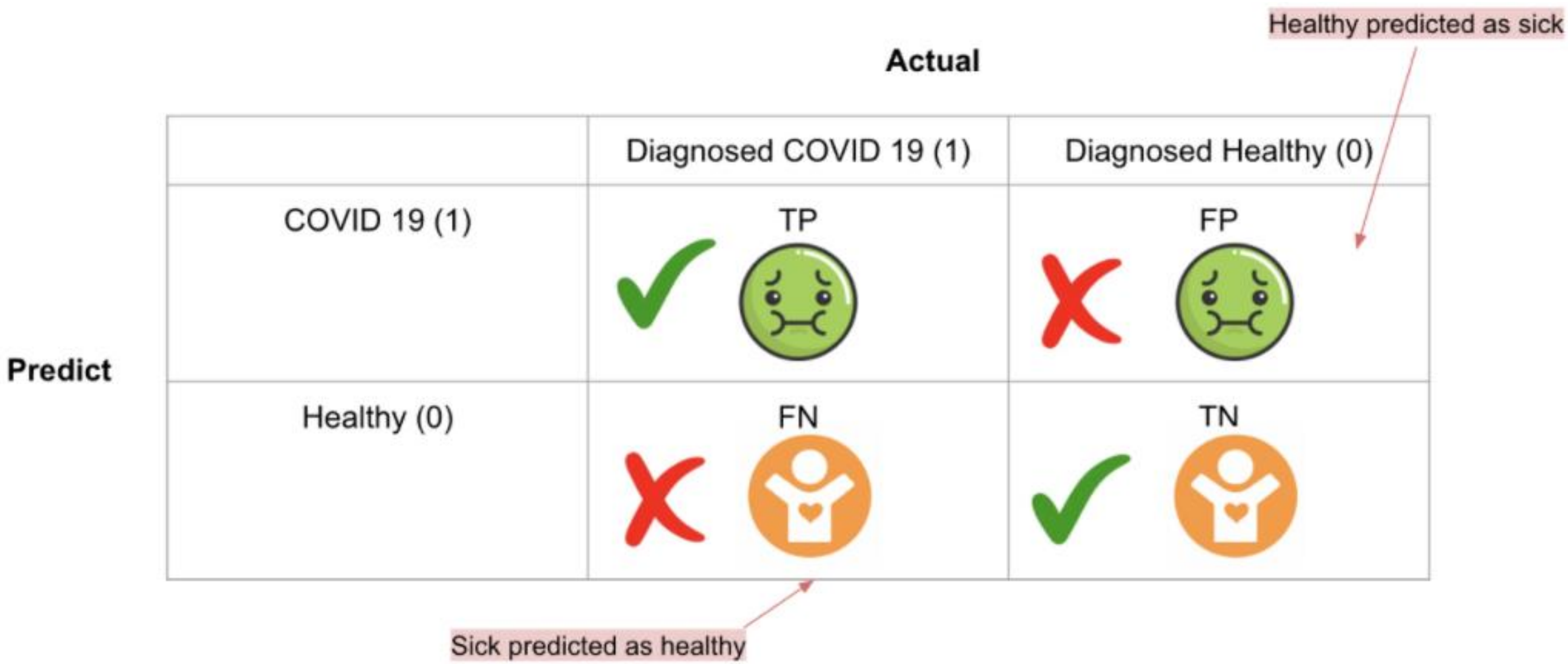
True Negative (TN) : Observation is negative, and is predicted to be negative.

False Positive (FP) : Observation is negative, but is predicted positive.

Evaluation of neural networks

Quality of results - classifications

- Case 1: **Cost of FN > Cost of FP**.



- If we predict COVID-19 patients as healthy people, they do not need to be quarantine.
- There would be a massive number of COVID-19 infections.
- The cost of FNs is much higher than the cost of FPs.

Quality of results - classifications

- Case 1: **Cost of FP > Cost of FN**.

Not spam predicted as spam

		Actual	
		Spam (1)	Not Spam (0)
Predict	Spam (1)	TP	FP
	Not Spam (0)	FN	TN

Spam predicted as not spam

- If we predict important emails as spam, they may be unreadable.
- If we predict spam mails as not spam, they may be read unnecessarily.
- The cost of FPs is much higher than the cost of FNs.

Quality of results - classifications

- Accuracy is the basic classification metric.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

- Accuracy is the proportion of true results among the total number of cases examined.
- Easily suited for binary as well as a multiclass classification problem.
- Accuracy is a valid choice of evaluation for classification problems which are well balanced.

Evaluation of neural networks

Quality of results - classifications

- Precision is the number of correct positive results divided by the number of positive results predicted by the classifier.

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

- Precision evaluation metric is a valid choice when we want to be very sure of our prediction.

Evaluation of neural networks

Quality of results - classifications

- Recall is the number of correct positive results divided by the number of all samples that should have been identified as positive.

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

- It is measure of positive examples labeled as positive by classifier.

Quality of results - classifications

- F1 score is best if we can get a single score that kind of represents both Precision and Recall.

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

- It is good choice when we want to have a model with both good precision and recall.
- F1 score sort of maintains a balance between the precision and recall for your classifier.
- The main problem with the F1 score is that it gives equal weight to precision and recall.

Quality of results – classifications

Example

Bad Loan = 1

Good Loan = 0

Cost of **FN** > Cost of **FP**

Actual

Predict		**Bad Loan (1)**	**Good Loan (0)**
	Bad Loan (1)	✓ TP - 559	✗ FP - 0
	Good Loan (0)	✗ FN - 33	✓ TN - 22

Accuracy: Out of the total prediction made, how many did we predict correctly?

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuray = (559+22)/(559+22+33+0) = 95%

Precision: Out of the loan that is predicted as a bad loan, how many did we classify correctly?

$$\textit{Precision} = \frac{TP}{TP + FP}$$

Precision = *559/(559+0) = 100%*

Recall: Out of the **actual** bad loan, how many did we correctly predict as a bad loan?

$$\textit{Recall} = \frac{TP}{TP + FN}$$

Instead of using **accuracy**,we should evaluate **recall**. If we can decrease **FN**, the recall will increase.

$\boxed{\textbf{\textit{Recall}} = 559/(559+33) = 94.5\%}$

Quality of results - clustering

- For clustering problems, the number of clusters should be small.
- Intracluster distances should be small and inter-cluster distances should be large.

$$E = \alpha(\text{No. of clusters}) + \beta \sum_{\substack{\text{all} \\ \text{clusters}}} (\text{Intra-cluster Distance})$$
$$- \gamma \sum_{\substack{\text{pairs of} \\ \text{clusters}}} (\text{Inter-cluster Distance})$$

α, β, γ are non-negative

Content

- Applications
  - Classification
  - Clustering
  - Pattern association
  - Function approximation
  - Forecasting
- Evaluation of neural networks
  - Quality of results
  - Generalizability
  - Computational resources

Evaluation of neural networks

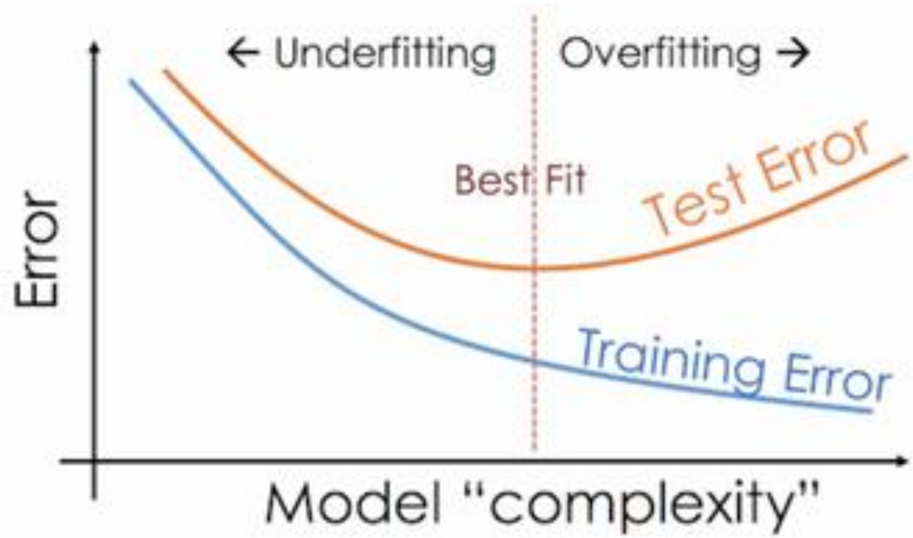
Generalizability

- A system may perform well on the training data.
- But, the system must perform well on new test data distinct from training data.
- Available data is separated into two parts, the training data and the test data.
- The training data set is used to train the model.
- The test set (unseen data for model) is used to test the model after training.
- Training set and test set should be disjoint sets.

Evaluation of neural networks

Generalizability

- In overtraining, the network performs well on training set but not test set.



- One way to avoid of overtraining is evaluation of the system using test data as learning proceeds.
- If there is a succession for only training data and not for the test data, overtraining is considered to have occurred.

Content

- Applications
  - Classification
  - Clustering
  - Pattern association
  - Function approximation
  - Forecasting
- Evaluation of neural networks
  - Quality of results
  - Generalizability
  - Computational resources

Evaluation of neural networks

Computational resources

- Once training is complete, many neural networks generally take up very little time in execution.
- Training the networks can take a very long time.
- Training time increases rapidly with the size of the networks and complexity of the problem.
- For fast training, the problem can be break down into smaller subproblems and solve each one separately.

Evaluation of neural networks

Computational resources

- The capabilities of a network are limited by its size.
- The use of large networks increases training time and reduces generalizability.
- Size of a network can be measured in terms of the number of nodes, connections, and layers in a network.
- Complexity of node functions contributes to network complexity measures.