CHAPTER 6

Unit VI

Multiclass Classification

Syllabus Topics

Multiclass Classification, Semi-Supervised Classification, Reinforcement learning, Systematic, Learning, Wholistic learning and multi-perspective learning. Metrics for Evaluating Classifier Performance : Accuracy, Error Rate, precision, Recall, Sensitivity, Specificity; Evaluating the Accuracy of a Classifier : Holdout Method, Random Sub sampling and Cross-Validation.

Syllabus Topic : Multiclass Classification

6.1 Multiclass Classification

6.1.1 Introduction to Multiclass Classification

- In Multiclass classification, there are N different classes.
- Each of the training point belongs to one of N different classes.
- The goal is to predict a class label to an Unknown tuple.

☞ Two Approaches in multiclass classification

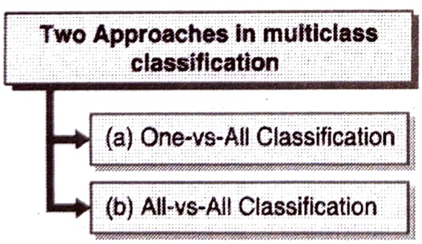

Fig. 6.1.1 : Two Approaches in multiclass classification

➜ (a) One-vs-All Classification

- Select a good technique for building a Binary Classifier (e.g. SVM).
- Build N different Binary classifiers.
- Classifier i is trained using tuples of class i as the positive class and the remaining tuples as the negative class.
- To classify an unknown tuple, X, the set of classifiers vote collectively.
- If classifier i predicts positive class for X, then class I gets one vote.
- If the classifier i predicts the negative class for X, then each of the classes except I gets one vote.
- The class with maximum votes is assigned to X.

➜ (b) All-vs-All Classification

- This is an alternative approach that learns a classifier for each pair of classes.
- Given N classes, construct n(n-1)/2 classifiers.
- A classifier is trained using tuples of the two classes.
- For classifying an unknown tuple X, each classifier votes.
- The class with maximum number of votes is assigned to the unknown tuple X.
- All-vs-All approach is better as compared to One-vs-All.

Syllabus Topic : Semi-Supervised Classification

6.2 Semi-Supervised Classification

- A semi-supervised classification uses labeled and unlabeled data to build a classifier.
- Following are the two forms of Semi-Supervised classification :

Semi-Supervised Classification

- (a) Self-training
- (b) Co-training

Fig. 6.2.1 : Semi-Supervised Classification

(a) Self-training

- It is one of the simplest form of semi-supervised classification.
- It first builds the classifier using the labeled data.
- Then the classifier tries to label the unlabeled data.
- The tuple with most confident label prediction is added to the labeled data.
- This process is repeated.
- One of the drawback of the method is that it may reinforce errors.

(b) Co-training

- This is another form of semi-supervised classification.
- In this approach two or more classifiers teach each other.
- Each learner uses different and independent set of features for each tuple.
- If the feature set is split into two sets and train two classifiers f1 and f2.
- Then f1 and f2 are used to predict the class labels for the unlabeled data.
- Each classifier then teaches the other in that tuple having the most confident prediction from f1 is added to the set of labeled data for f2(along with its label).
- The tuple having the most confident prediction from f2 is added to the set of labeled data for f1.
- Cotraining is less error prone as compared to self training.

Syllabus Topic : Reinforcement Learning

6.3 Reinforcement Learning

(SPPU - Dec. 16, May 17)

Q. Briefly explain the reinforcement learning.

Dec. 16, May 17, 6 Marks

6.3.1 Introduction to Reinforcement

- Reinforcement learning is based on goal-directed learning from interaction.
- Reinforcement learning maximizes a numerical reward signal by mapping the situations to actions.
- In machine learning learner knows which action to take but in reinforcement learning, learner doesn't know the action, it discover which action gives most reward signal.
- Reinforcement learning is by characterizing a learning problem not by method.
- Reinforcement learning is different from supervised learning, as alone it is not adequate for learning from interaction.
- In this case, the agent has to act and get the learning through experience.
- All reinforcement learning agents have explicit goals and are intelligence to find the aspects of their environments, so accordingly they can select the actions to control their environments.

Example

- In a chess game, player makes a move based on planning of move, expecting possible replies and even counter replies. Then player takes immediate and spontaneous judgment and plays the move.
- In this example the agent (player) uses its experience to improve its performance and evaluate positions to improve his play over the time.

6.3.2 Elements of Reinforcement Learning

Main sub-elements of a reinforcement learning system are :

Elements of Reinforcement Learning

- 1. A policy
- 2. A reward function
- 3. A value function
- 4. A model of the environment (optional)

Fig. 6.3.1 : Elements of Reinforcement Learning

→ **1. A policy**

The learning agent's manner of behaving at a given time.

→ **2. A reward function**

The purpose in a reinforcement learning problem.

→ **3. A value function**

What is good over the future or in the long run?

→ **4. A model of the environment (optional)**

It is used for planning and predict the resultant next state and next reward.

6.3.3 Reinforcement Function and Environment Function

- It uses knowledge acquired so far and while exploring, the action leads to learning through either rewards or penalties.
- Rewards are related to specific actions and value function is the collective effect.
- To get the correct responses, environment needs to be model so that it can accept the inputs from changing scenarios and finally can produce the optimized value.
- Fig 6.3.2 shows the typical reinforcement-learning scenario where action lead to reward.
- So for every action, there are environment as well as reinforcement functions.

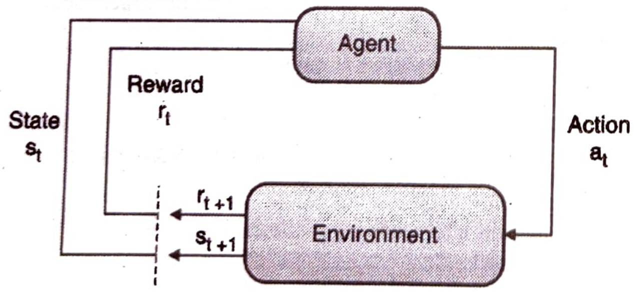

Fig. 6.3.2 : Reinforcement-learning scenario

Syllabus Topic : Wholistic Learning

6.3.4 Whole System Learning

→ **(SPPU - Dec. 16)**

Q. What is meant by whole system learning ?

Dec. 16, 4 Marks

- Systematic learning considers complete system, its subsystem and the interactions between the systems for learning. Based on this it makes the decisions.
- It builds the systematic information which is useful for analysis.
- Systematic learning is interactive and driven by environment which is specific to the problem.

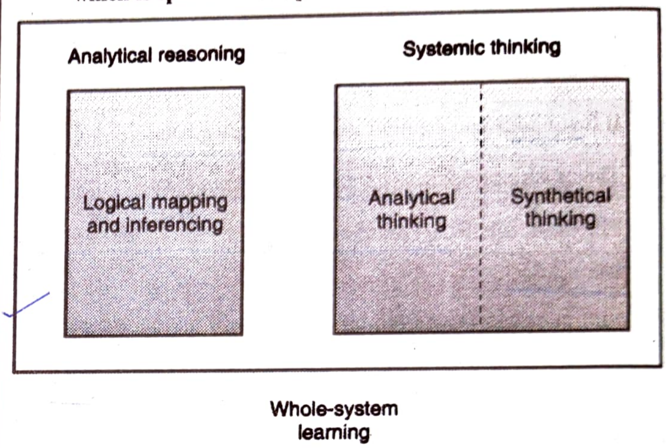

Fig. 6.3.3 : Whole-System learning

Syllabus Topic : Systematic Learning

6.4 Systematic Learning

- Systemic learning is about understanding systems, subsystems, and systemic impact of various actions, decisions within the system, and decisions in a systemic environment.
- It is more about learning about the actions and interactions from a systemic perspective.
- It identifies a system and builds systemic information.
- Building of this information is done from the analysis of perspectives with reference to systemic impact.
- This learning includes multiple perspectives and collection of data from all parts of system.
- It also includes data and decision analysis related to impact.
- Decisions are part of a learning process, and the learning takes place with every decision and its outcome.
- The knowledge augmentation takes place with every decision and decision-based learning.

- This learning is interactive and driven by environment, which include different parts of the system.
- The system dependency of learning is controlled and is specific to the problem and system.

Syllabus Topic : Multi-Perspective Learning

6.5 Multi-Perspective Decision Making for Big Data and Multi-Perspective Learning for Big Data

→ (SPPU - Dec. 15, May 16, Dec. 16, May 17, Dec. 17)

Q. Write a note on multi-perspective learning. **Dec. 15, May 17, Dec. 17, 4 Marks**

Q. Write short note on multi-perspective decision making. **May 16, 6 Marks**

Q. What is meant by multi perspective decision making? Explain. **Dec. 16, 6 Marks**

6.5.1 Fundamental of Multi-perspective Decision Making and Multi-perspective Learning

- Multi-perspective Learning is needed for Multi-perspective Decision making
- Multi-perspective Learning refers to learning from knowledge and information collected from different perspectives.
- Multi-perspective Learning builds knowledge from various perspectives so that it can be used for decision making process.

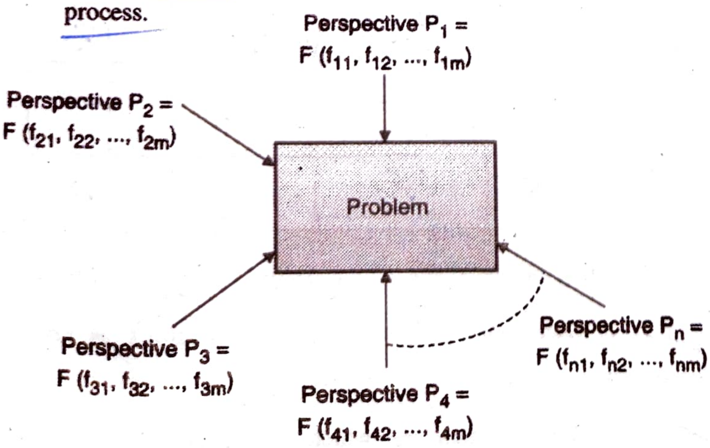

Fig. 6.5.1 : Multi-perspective Learning

- The perspective includes context, scenario and situation, the way we look at a particular decision problem.
- In Fig. 6.5.1, P_1, P_2, P_3..., P_n refers to different perspective in the Learning process.
- Each of this perspective is represented as a function of features.
- There may be an overlap among the perspectives.
- Feature difference may be there as some features which possibly visible form one perspective may not be visible from the other perspective.
- The representative feature set should contain all the possible features.

6.5.2 Influence Diagram

- Perspective based information can be represented as an influence diagram.
- Helps in getting the context of the decision making
- It is a graphical representation of the decision situation
- It shows the relationships among objects and actions.
- These relationships may be mapped to probabilities.
- The Fig. 6.5.2 represents an influence diagram for a market scenario and relationship between marketing and budget, product, price, cost and profit.

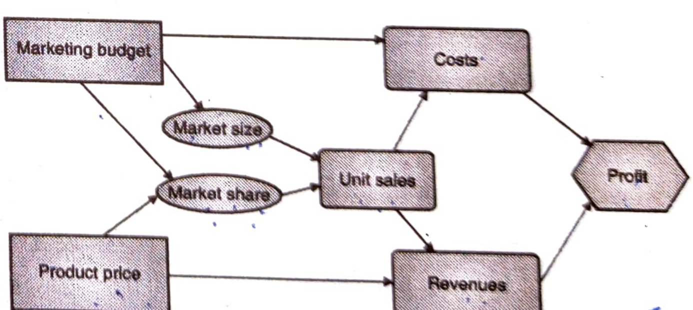

Fig. 6.5.2 : Influence Diagram

- The relationships may also be represented using a decision tree is shown in Fig. 6.5.3.
- Based on parameters measurements, the decision path of decision tree is decided.
- Decision rules can be represented on a decision tree.

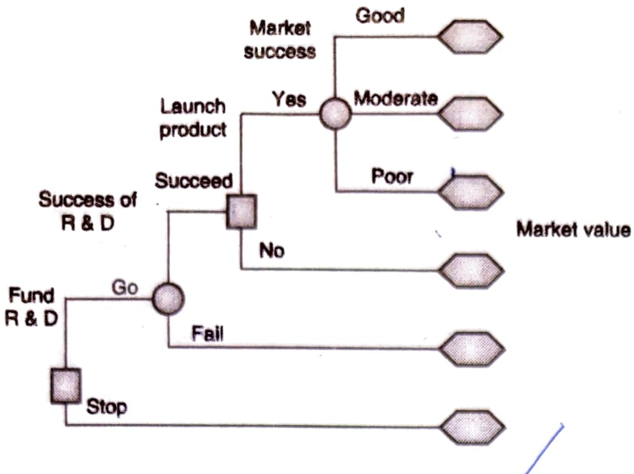

Fig. 6.5.3 : Decision tree

Syllabus Topic : Metrics for Evaluating Classifier Performance : Accuracy, Error Rate, Precision, Recall, Sensitivity, Specificity

6.6 Model Evaluation and Selection

- Validation test data is very useful to estimate the accuracy of model.
- Various methods for estimating a classifier's accuracy are given below. All of them are based on randomly sampled partitions of data :
  - o Holdout method
  - o Random subsampling
  - o Cross-validation
  - o Bootstrap
- If we want to compare classifiers to select the best one then the following methods are used :
  - o Confidence intervals
  - o Cost-benefit analysis and Receiver Operating Characteristic (ROC) Curves

6.6.1 Accuracy and Error Measures

Accuracy of a classifier M, acc(M) is the percentage of test set tuples that are correctly classified by the model M.

Basic concepts

1. Partition the data randomly into three sets

- Trainingset, validation set and test set.
- Training set is the subset of data used to train/build the model.
- Test set is a set of instances that have not been used in the training process. The model's performance is evaluated on unseen data. Testing just estimates the probability of success on unknown data.
- Validation data is used for parameter tuning but it cannot be the test data. Validation data can be the training data, or a subset of training data.
- Generalization Error : Model error on the test data.

2. Success

Instance (record) class is predicted correctly.

3. Error

Instance class is predicted incorrectly.

4. The confusion matrix

- It is a useful tool for analyzing how well your classifier can recognize tuples of different classes.
- If we have only two way classification then only four classification outcomes are possible which are given below in the form of a confusion matrix :

		Predicted class		
	Class Label	C_1 P	C_2 N	Total
Actual class	C_1	True Positives (TP)	False Negatives (FN)	P
	C_2	False Positives (FP)	True Negatives (TN)	N
	Total	P'	N'	All

- **TP** : Class members which are classified as class members.
- **TN** : Class non-members which are classified as non-members.
- **FP** : Class non-members which are classified as class members.

- FN : Class members which are classified as class non-members.
- P : Number of positive tuples.
- N : The number of negative tuples.
- P' : The number of tuples that were labeled as positive.
- N' : The number of tuples that were labeled as negative
- All : Total number of tuple i.e. TP + FN + FP + TN or P + N or P' + N'

5. **Sensitivity** : True Positive recognition rate which is the proportion of positive tuples that are correctly identified

$$\text{Sensitivity} = TP/P$$

6. **Specificity** : True Negative recognition rate which is the proportion of negative tuples that are correctly identified

$$\text{Specificity} = TN/N$$

7. **Classifier accuracy or recognition rate** : Percentage of test set tuples that are correctly classified

$$\text{Accuracy} = (TP + TN)/All$$

OR

$$\text{Accuracy} = \frac{TP + TN}{P + N}$$

Accuracy is also a function of sensitivity and specificity :

$$\text{Accuracy} = \text{Sensitivity}\frac{P}{(P+N)} + \text{Specificity}\frac{N}{(P+N)}$$

8. **Error rate** : A percentage of errors made over the whole set of instances (records) used for testing.

$$\text{Error rate} = 1 - \text{accuracy, or Error rate} = (FP + FN)/All$$

Or

$$\text{Error rate} = \frac{FP + FN}{P + N}$$

9. **Precision** : Percentage of tuples which are correctly classified as positive are actual positive. It is a measure of exactness.

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

10. **Recall** : Percentage of positive tuples which the classifier labelled as positive. It is a measure of completeness.

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

11. **Fmeasure (F_1orF-score)** : Harmonic mean of precision and recall,

$$F = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

12. **F_β** : Weighted measure of precision and recall and assigns β times as much weight to recall as to precision

$$F_\beta = \frac{(1+\beta^2) \times \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}}$$

where β is a non-negative real number.

13. **Classifiers can also be compared with respect to**

- Speed
- Robustness
- Scalability
- Interpretability

14. **Re-substitution error rate**

- Re-substitution error rate is a performance measure and is equivalent to training data error rate.
- It is difficult to get 0% error rate but it can be minimized, so low error rate is always preferable.

Syllabus Topic : Evaluating the Accuracy of a Classifier : Holdout Method

6.6.2 Holdout

- In holdout method, data is divided into training data set and testing data set (usually 1/3 for testing, 2/3 for training).

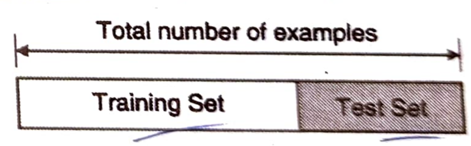

Fig. 6.6.1

- To train the classifier, training data set is used and once the classifier is constructed then use test data set to estimate the error rate of the classifier.
- If the training is more than better model is constructed and if the test data is more than more accurate the error estimates.
- **Problem** : The samples might not be representative. For example, some classes might be represented with very few instances or even with no instances at all.
- **Solution** : stratification is the method which ensures that both training and testing data have equal number of samples of same class.

Syllabus Topic : Random Sub-sampling

6.6.3 Random Sub-sampling

- It is a variation of the holdout method.
- The holdout method is repeated k times.
- Each split randomly selects a fixed number example without replacement.

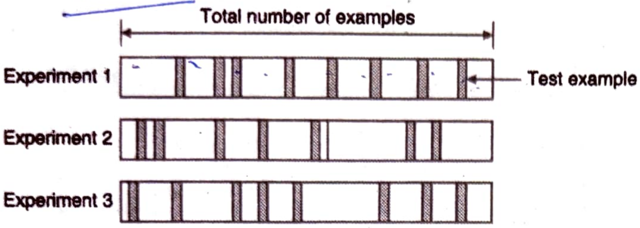

Fig. 6.6.2

- For each data split we retrain the classifier from scratch with the training examples and estimate E_i with the test examples.
- The overall accuracy is calculated by taking the average of the accuracies obtained from each iteration.

$$E = \frac{1}{K}\sum_{i=1}^{K} E_i$$

$$E = \frac{1}{K}\sum_{i=1}^{K} E_i$$

Syllabus Topic : Cross-Validation (CV)

6.6.4 Cross-Validation (CV)

Avoids overlapping test sets.

☞ **k-fold cross-validation**

- **First step :** Data is split into k subsets of equal size (usually by random sampling).

3

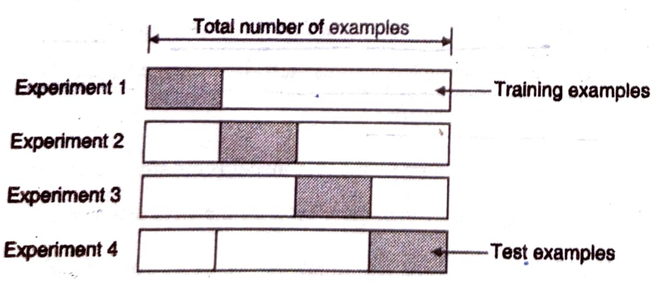

Fig. 6.6.3

- **Second step :** Each subset in turn is used for testing and the remainder for training.
- The advantage is that all the examples are used for both training and testing.
- The error estimates are averaged to yield an overall error estimate.

$$E = \frac{1}{K}\sum_{i=1}^{K} E_i$$

☞ **Leave-one-out cross validation**

- If dataset has N examples, then N experiments to be performed for Leave-one-out cross validation.
- For every experiment, training uses N-1 examples and remaining example for testing.
- The average error rate on test examples gives the true error.

$$E = \frac{1}{N}\sum_{i=1}^{N} E_i$$

- **Stratified cross-validation :** Subsets are stratified before the cross-validation is performed.

☞ **Stratified ten-fold cross-validation**

- This gives accurate estimate of evaluation.
- The estimate's variance get reduced due to stratification.
- Ten-fold cross-validation is repeated ten times and finally the results are averaged based on the previous 10 results.

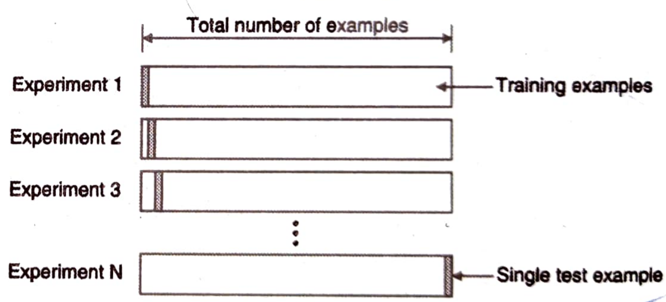

Fig. 6.6.4

6.7 Solved University Questions and Answers

Q. 1 For each of the following queries, identify and write the type of data mining task.

i) Find all credit applicants who are poor credit risks.

ii) Identify customers with similar buying habits.

iii) Find all items which are frequently purchased with milk. **(Dec. 15, 6 Marks)**

Ans. :

i) **Classification :** Based on credit applications , customers can be classified in various classes like poor, medium and high credit risk types of customers

ii) **Clustering :** Clusters can be formed based on similar type of buying patterns. Then the customers belongs to those clusters can be identified.

iii) **Association :** Various items which has been frequently purchased with milk can be identified with association data mining task. Based on the support and confidence, milk can be associated with those frequent items.

Q. 2 Consider the ten records given below : **(Dec. 15, 8 Marks)**

ID	Income	Credit	Class	X_i
1	4	Excellent	h_1	X_4
2	3	Good	h_1	X_7
3	2	Excellent	h_1	X_2
4	3	Good	h_1	X_7
5	4	Good	h_1	X_8
6	2	Excellent	h_1	X_2
7	3	Bad	h_2	X_{11}
8	2	Bad	h_2	X_{10}
9	3	Bad	h_3	X_{11}
10	1	Bad	h_4	X_9

Calculate the prior probabilities of each of the class h_1, h_2, h_3, h_4 and probabilities for data points X_1, X_4, X_7 and X_8, belonging to the class h_1.

Ans. :

Assign ten data values for all combinations of credit and income :

	1	2	3
Excellent	x_1	x_3	x_6
Good	x_2	x_4	x_5
Bad	x_7	$x_{8/9}$	x_{10}

From training data

$$P(h_1) = \frac{6}{10} = 60\ \%$$

$$P(h_2) = \frac{2}{10} = 20\ \%$$

$$P(h_3) = \frac{1}{10} = 10\ \%$$

$$P(h_4) = \frac{1}{10} = 10\ \%$$

Q. 3 What are similarities and differences between reinforcement learning and artificial intelligence algorithms ? **(Dec. 16, 5 Marks)**

Ans. :

Reinforcement Learning	Artificial Intelligence
Reinforcement learning is a branch of Artificial Intelligence and type of machine learning algorithms.	Artificial Intelligence (AI) is an area in computer science that emphasizes the creation of intelligent machines.
To maximize its performance, it allows software agents and machines to automatically determine the ideal behavior within a specific context.	One of the fundamental building blocks of artificial intelligence (AI) solutions is Learning. From a conceptual standpoint, learning is a process that improves the knowledge of an AI program by making observations about its environment.
The feedback received from the environment is used by the machine or software agent to learn its behavior.	Based on the feedback characteristics, AI Learning models can be classified into supervised, unsupervised, semi-supervised or reinforced.
Applications : Manufacturing, inventory management, delivery management, finance sector.	**Applications :** Expert systems, Speech recognition and Machine vision, Natural Language processing.

Similarity between reinforcement learning and systematic machine learning :

The similarity between reinforcement learning and systematic machine learning is both are Adapts to evolving environment.

❑❑❑