Machine Learning Methods

Different Approaches

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Machine Learning methods

Classification and Prediction

- Supervised Method
- Requires Labelled data
- Various Performance indicators
- Most common Performance measure ??

What do you mean by labelled data

- Data on student placements
- Data on student placements categorized Low medium high package range
- Customer Reviews
- MRI images of brain tumour

Classification – The Process

1. Training and Test Data
2. Features
3. Algorithm
4. Evaluation and refinements (Error analysis)
5. Real time prediction
6. Regression – Predicting continuous values

Supervised vs. Unsupervised Learning

- Supervised learning (classification)
  - Supervision: The training data (observations, measurements, etc.) are accompanied by labels indicating the class of the observations
  - New data is classified based on the training set
- Unsupervised learning (clustering)
  - The class labels of training data is unknown
  - Given a set of measurements, observations, etc. with the aim of establishing the existence of classes or clusters in the data

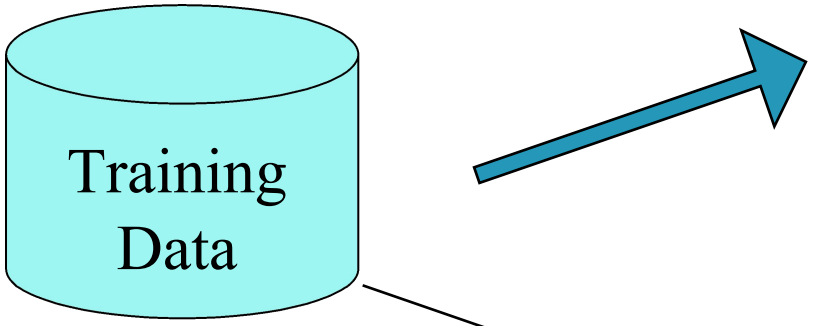
Classification vs. Prediction

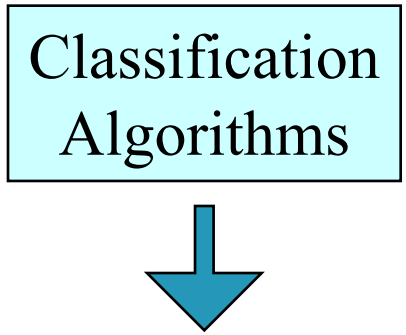
- Classification
  - predicts categorical class labels (discrete or nominal)
  - classifies data (constructs a model) based on the training set and the values (class labels) in a classifying attribute and uses it in classifying new data
- Prediction
  - models continuous-valued functions, i.e., predicts unknown or missing values
- Typical applications
  - Credit/loan approval:
  - Medical diagnosis: if a tumor is cancerous or benign
  - Fraud detection: if a transaction is fraudulent
  - Web page categorization: which category it is

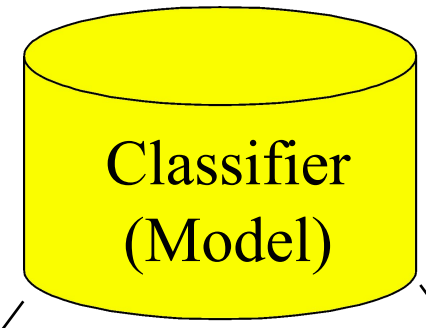
Classification—A Two-Step Process

- Model construction: describing a set of predetermined classes
  - Each tuple/sample is assumed to belong to a predefined class, as determined by the class label attribute
  - The set of tuples used for model construction is training set
  - The model is represented as classification rules, decision trees, or mathematical formulae
- Model usage: for classifying future or unknown objects
  - Estimate accuracy of the model
    - The known label of test sample is compared with the classified result from the model
    - Accuracy rate is the percentage of test set samples that are correctly classified by the model
    - Test set is independent of training set, otherwise over-fitting will occur
  - If the accuracy is acceptable, use the model to classify data tuples whose class labels are not known

Process (1): Model Construction



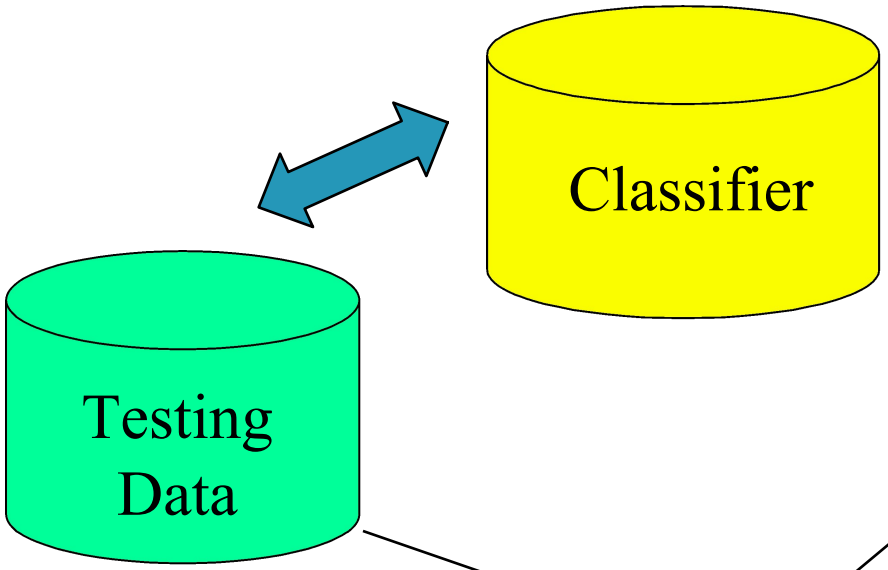




NAME	RANK	YEARS	TENURED
Mike	Assistant Prof	3	no
Mary	Assistant Prof	7	yes
Bill	Professor	2	yes
Jim	Associate Prof	7	yes
Dave	Assistant Prof	6	no
Anne	Associate Prof	3	no

IF rank = ‘professor’
OR years > 6
THEN tenured = ‘yes’

Process (2): Using the Model in Prediction



NAME	RANK	YEARS	TENURED
Tom	Assistant Prof	2	no
Merlisa	Associate Prof	7	no
George	Professor	5	yes
Joseph	Assistant Prof	7	yes

(Jeff, Professor, 4)

Tenured?

Yes

Supervised Learning

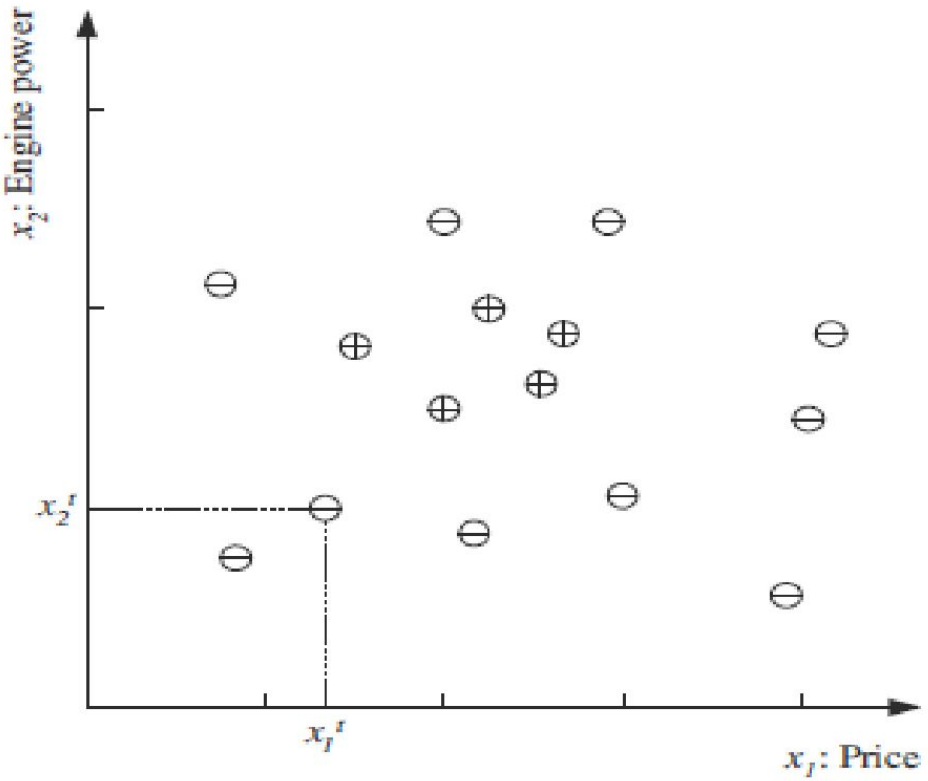
$Y=\{x^t,r^t\}$ for t=1….N

Model/rules/patterns to give value of result

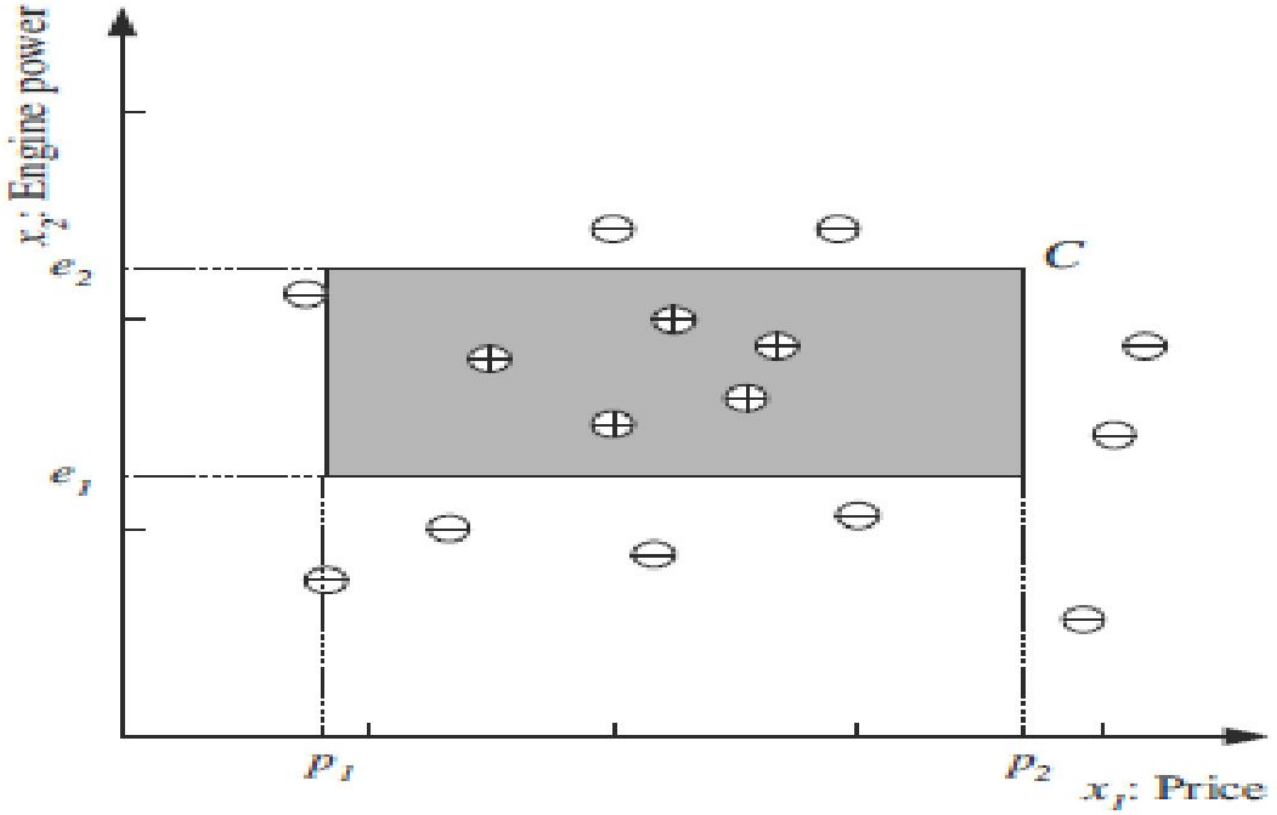
X^t is set of input variables

R^t is the result that will be predicted

Predicting if a car is a family car



Rectangle can differentiate between classes



Classification - Rule

- One possible rule

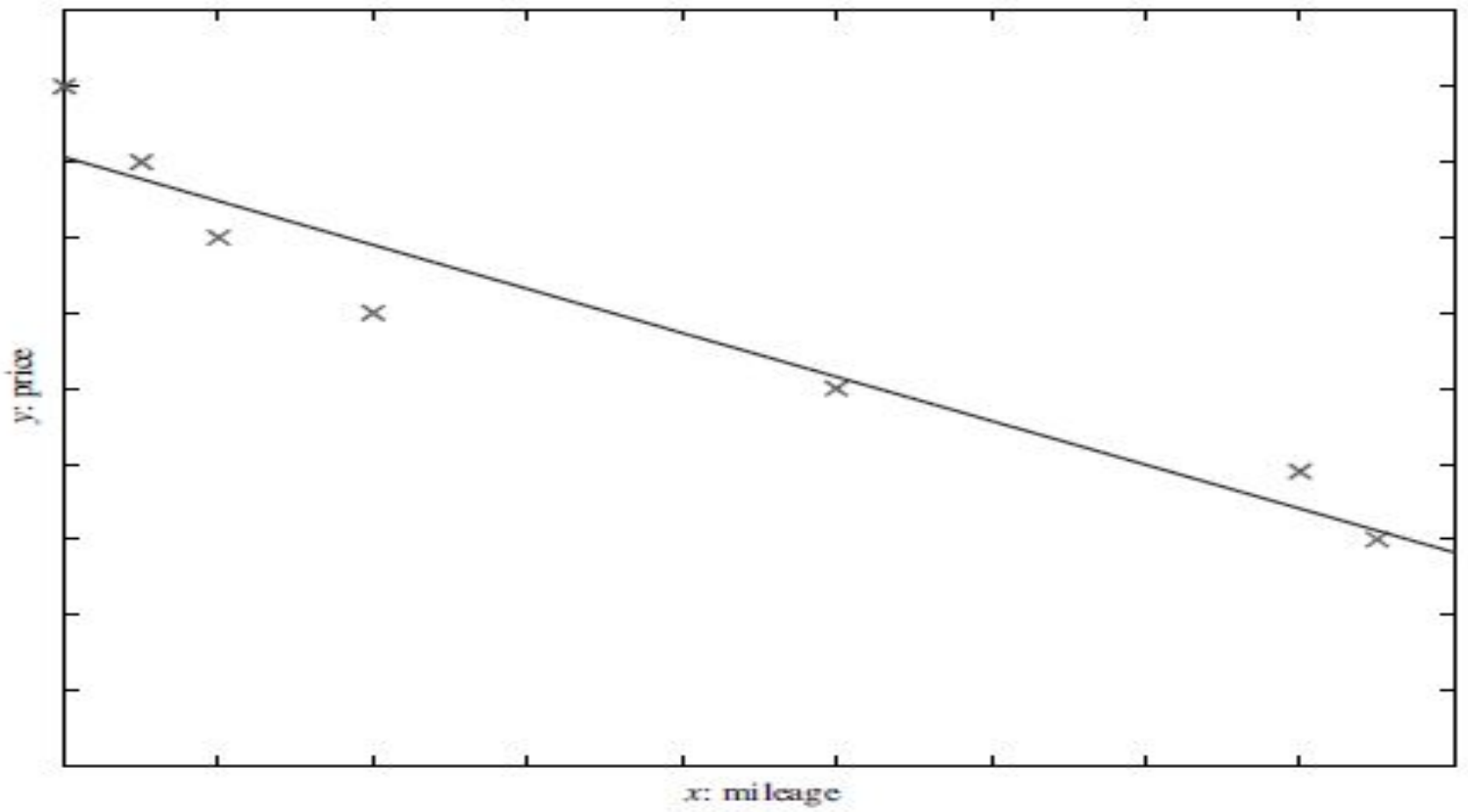
If (price>p1 AND price<p2) AND if(engine power> e1 AND if(engine power < e2)

- There can be many Rules

Regression

- Predicting Numeric value
  Examples
  Predicting rain/temperature
  Forecasting sales
  Predicting share Prices

Predictitng Price of a used car



Error

- Mismatch (Predicted y different from actual Y)
- Deviation (Predicted Y different from actual Y by how much)
- Confusion matrix
-

	A	B	C
A	Correct		
B		Correct	
C			Correct