Deeplearners@GMRIT

In association with leadingIndia.AI

Day #3

GMRIT Training Tomorrow's Engineers Today

GMR

Fundamentals of AI & Machine Learning

Dr V Prasad
Associate Professor,
Dept of CSE,

GMRIT
Training Tomorrow's Engineers Today

GMR

Todays Agenda

- Introduction to AI
- Difference between Learning Algorithms and Normal Algorithms
- Where we can apply learning algorithms
- Learning Techniques
- Disadvantages of Rule Based Systems
- Traditional AI Vs ML
- ML Techniques
- Prediction with Linear Regression

Questions to Deep Learners
in
E-Class Room
Intelligent Will Answer in Seconds

If they are 6 apples and if you take away 4, how many do you have?

Current Trend Setters Word-- "Google Says AI is used for removal of Blindness"

Artificial Intelligence is a way of **making a computer, a computer-controlled robot, or a software think intelligently**, in the similar manner the intelligent humans think.

Goals of AI

To Create Expert Systems

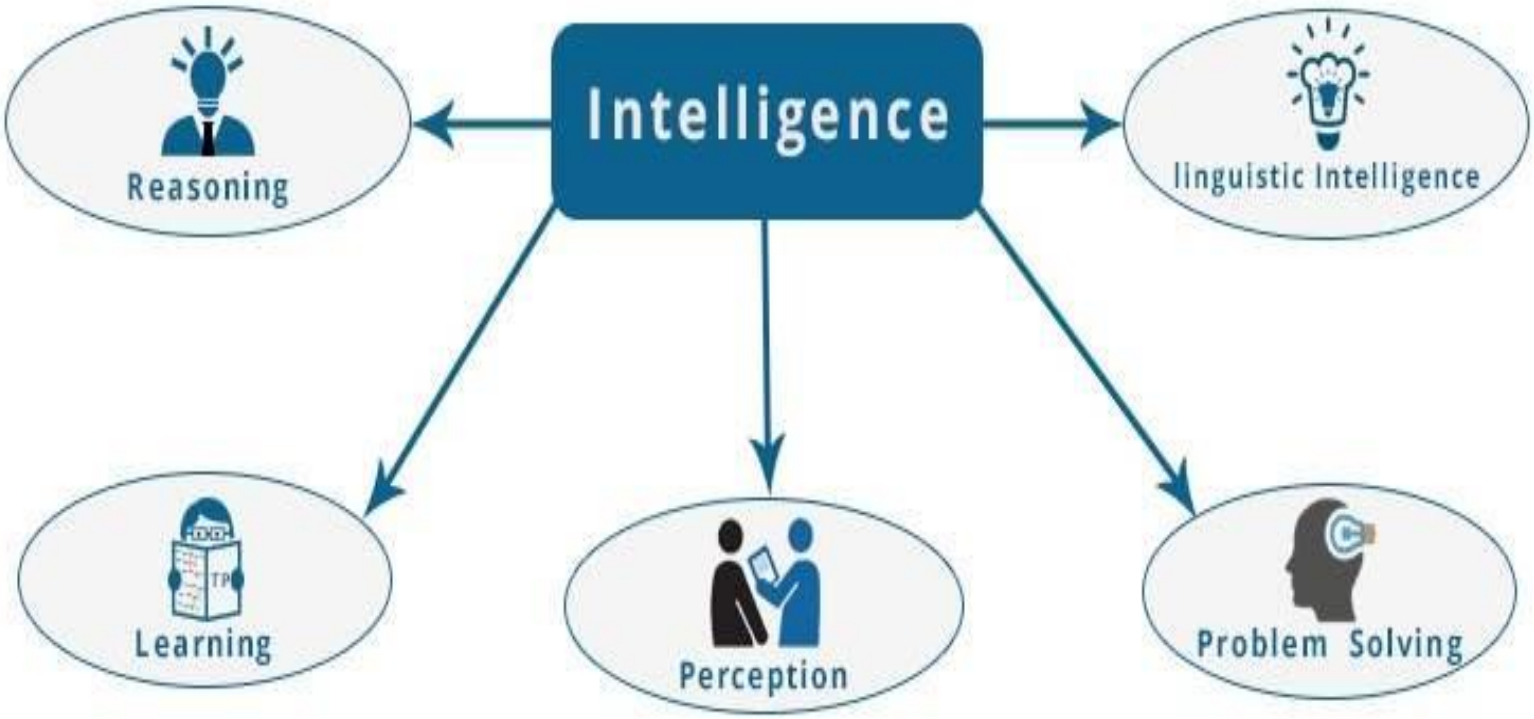
To Implement Human Intelligence in Machines

Artificial Intelligence , the two letter word sounds very BIG

Where Artificial means Simulation

& Intelligence means Ability of Thinking, Learning & Understanding.

Human-like Intelligence



Origin of AI

Alan Turing(1951)

Early enthusiasm 52-59:

Puzzle solving with the General Problem Solver

Mc. Carthy (1956) Machine can do things better than Human.

Machines are to be developed in such a way that they can
Think Intelligently & Act Intelligently

Russsell & Norvig:

- Acts like human (Turing test)
- Thinks like human (human-like patterns of thinking steps)
- Acts or thinks rationally (logically, correctly)

Some problems used to be thought of as AI

e. g., compiling Fortran in 1955,

- symbolic mathematics in 1965,
- pattern recognition in 1970,

Knowledge-based systems (expert systems) 1969-1979:

Ed Feigenbaum (Stanford): Knowledge is power!

MYCIN: Diagnosis of blood infections

AI becomes an industry:

Configuring computers & Robotic vision applications

AI turns more scientific, relies on more mathematically sophisticated tools:

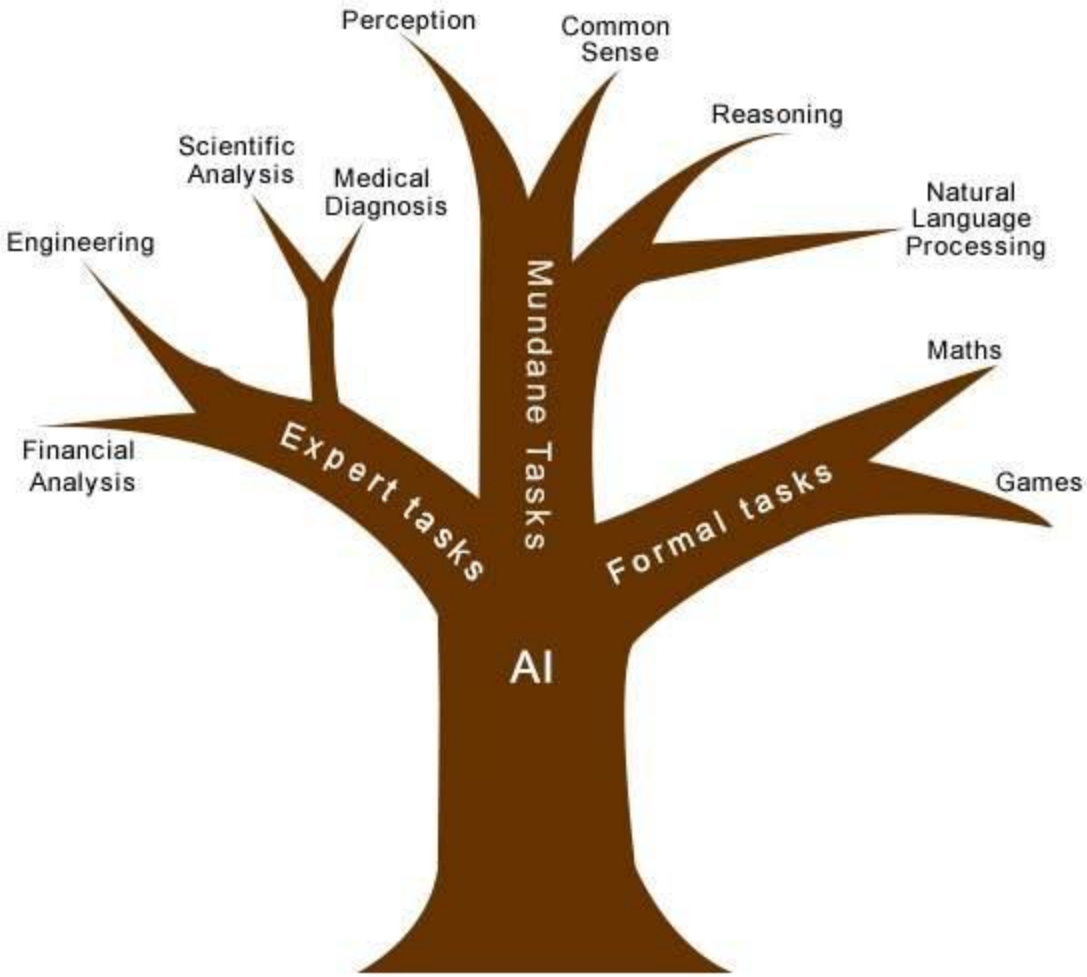
- Hidden Markov models
- Belief networks

Focus turns to building useful artifacts as opposed to solving the grand AI problem.

Turing Test

Alan Turing(1950) introduced Turing Test for evaluation of intelligence and published Computing Machinery and Intelligence.

AI Problems & Tasks



Real Life Applications of AI Research Areas

Expert Systems

Examples − Flight-tracking systems, Clinical systems.

Natural Language Processing

Examples: Google Now feature, speech recognition, Automatic voice output.

Neural Networks

Examples − Pattern recognition systems such as face recognition, character recognition, handwriting recognition.

Robotics

Examples − Industrial robots for moving, spraying, painting, precision checking, drilling, cleaning, coating, carving, etc.

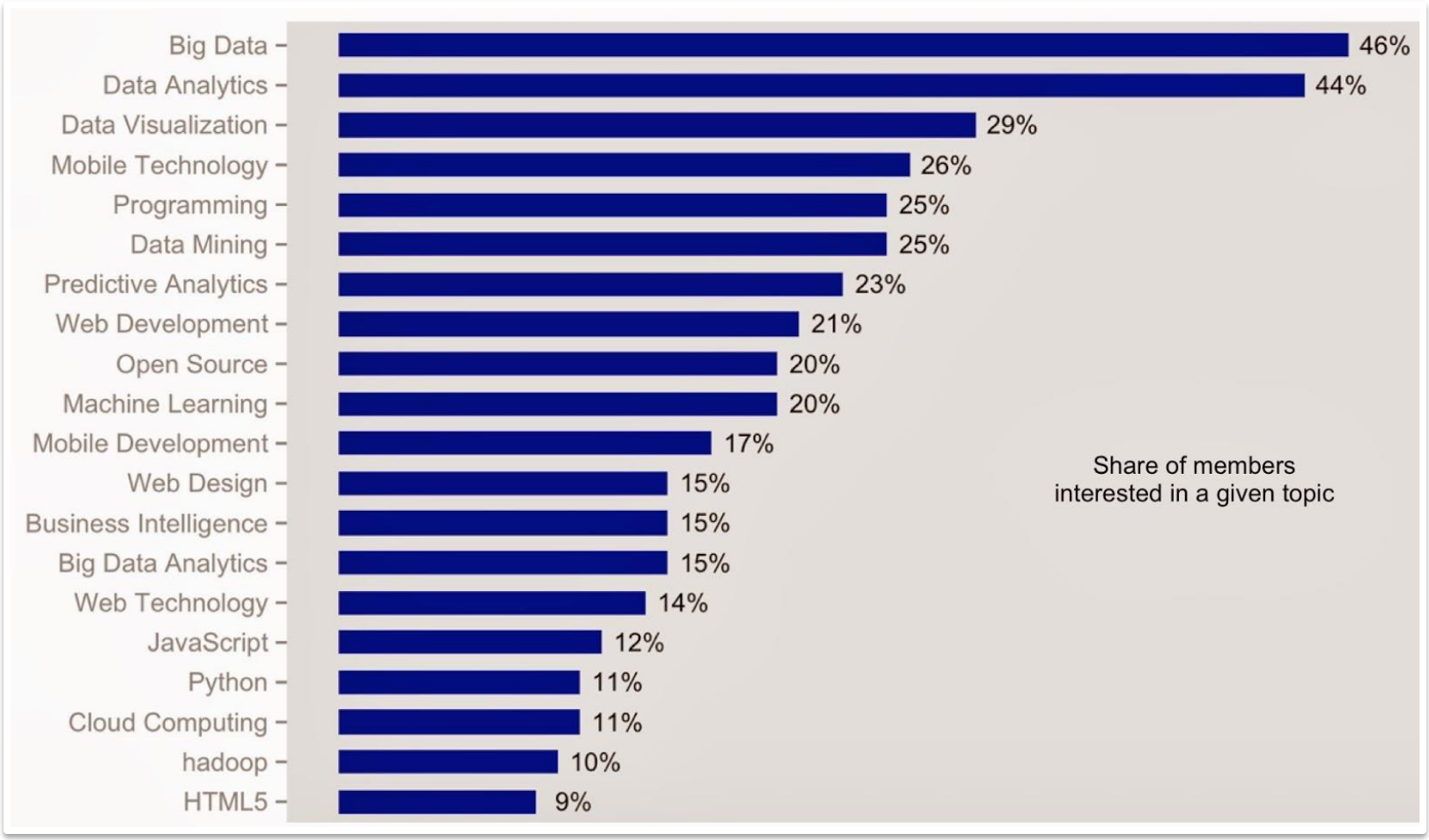
Fuzzy Logic Systems

Examples − Consumer electronics, automobiles, etc.

Questions to Deep Learners in E-Class Room Intelligent Will Answer in Seconds

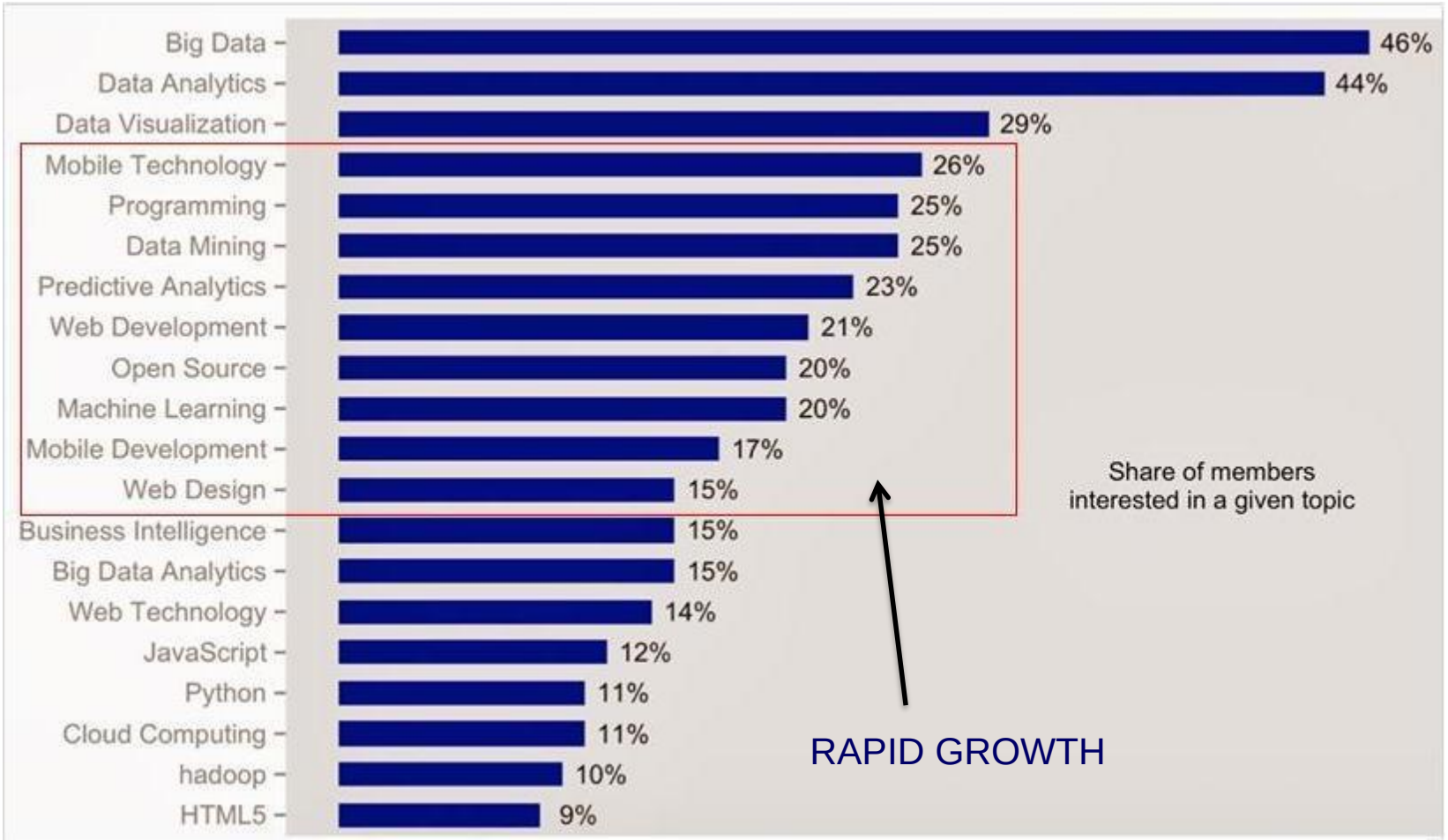
A is the father of B. But, B is not the son of A. How it is possible ?

General Statistics



SURVEY REPORT from Social Networks

Trends growing Fast In



Machine Learning Algorithms

- Machine Learning is a subfield of Computer Science where there is a study based on Data.
- Machine Learning has a strong tie with AI & Optimization .
- Machine Learning is used to Analyze Datasets.
- Problems related to Machine Learning
  - Clustering
  - Classification & Regression
  - Dimensionality Reduction
  - Prediction
  - Anomaly Detections
  - Neural Nets

- Basic Applications of Machine Learning are
  - SPAM FILTERING
  - OPTICAL CHARACTER RECOGNITION
  - SEARCH ENGINES
  - COMPUTER VISIONS

➔ Analyzing the Data is named as Simplification & Classification.

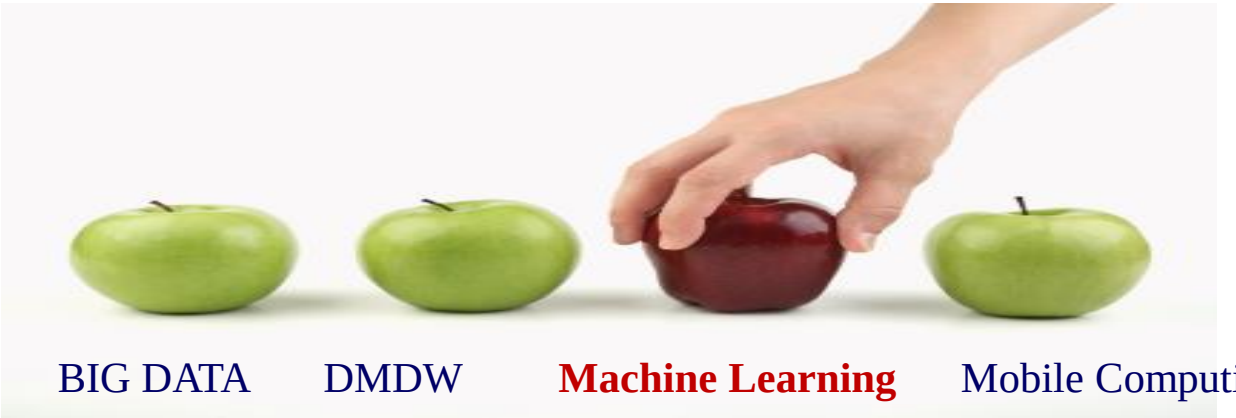
- Machine Learning are composed of two components namely

(a) Training Phase (Labeled Data)

Use some algorithms and encode training examples which are obtained from previous experiences.

(b) Prediction Phase(Classification through Unlabeled Data)

This is used to predict that of the class which is a new instance obtained , not exactly relevant with the trained data.



- Accuracy Levels are High in prediction when the data is consistent .
- Simplification and Classifications of Data Sets can be done through patterns like images / text /audio/video clips etc.....
- Deals with Missed Data.
- Noisy Data (Uncertainty and Errors).
- Transparency of Diagnostic Knowledge.
- Reduction of Tests.

Questions to Deep Learners in E-Class Room

Lets construct a simple static dataset and classify it accordingly.

Expert Systems

- Computer Professionals are able to provide Expert Systems to diagnose different kinds of diseases with High Accuracy.
- These Systems are based on Artificial Intelligence which helps the physician/consultant to minimize the costs and time and become expert in effective diagnoses.
- Medical Diagnose is implemented for Diseases related to Cancer , Diabetes, Heart Diseases , Lungs , Liver and so on……
- Recently and Android Mobile is released into the Market for Testing which consists of 121 Medical Tests based on the Symptoms which it identifies through persons mass/skin /saliva and so on …
- Some Examples for Android Apps developed are Disease Dictionary , Drugs Dictionary ,Feed Baby Pro , Pregnancy+,Muscle Triquer Pointer and so on …..

- No Intelligence without Learning .
- Predicting .
- Intelligent Data Analysis in AI from ML perspective can be in

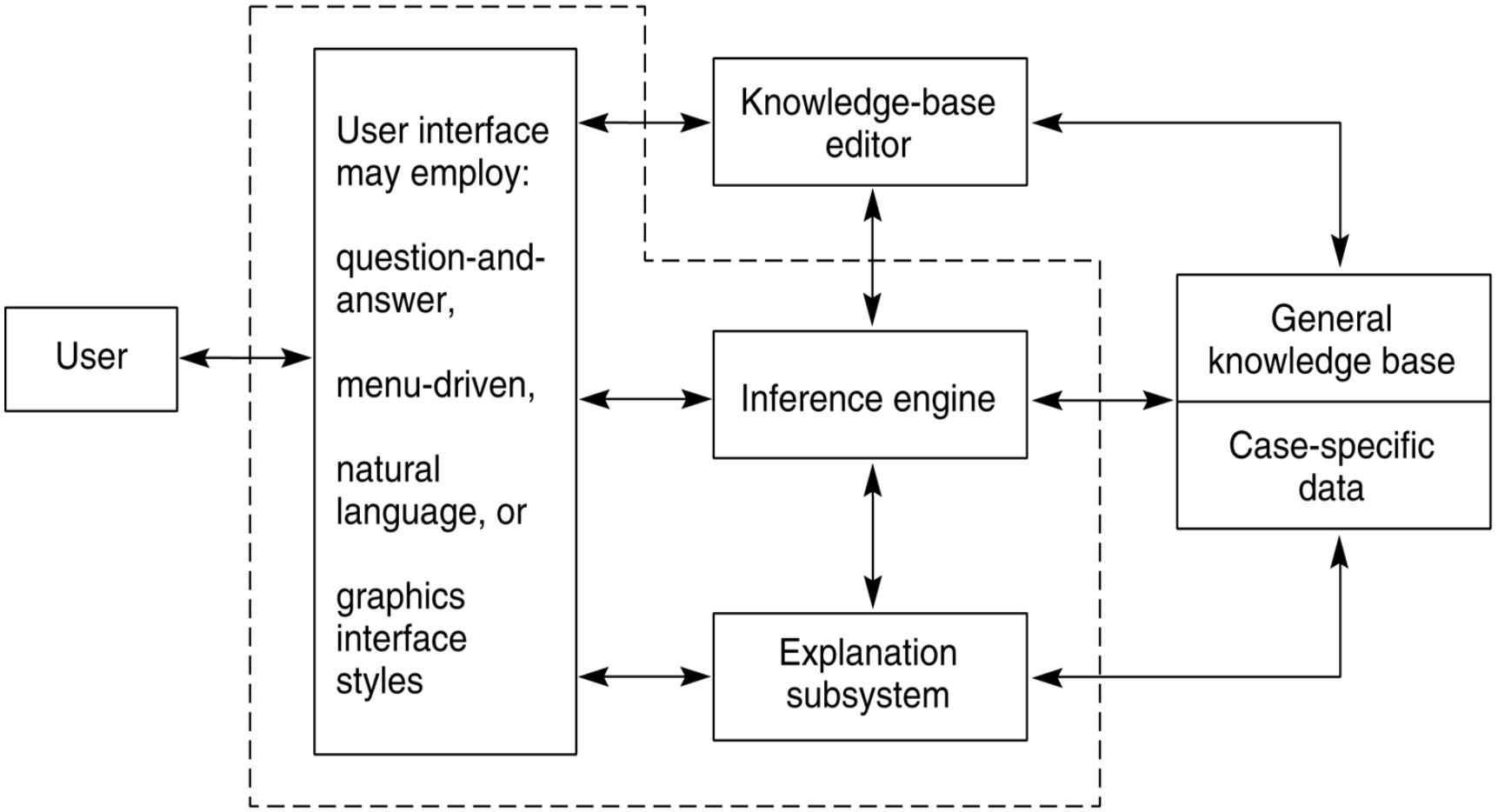
  (a) Historical Data

  (b) State of View

  (C) View on future exp

Traditional System/ Model Issues

- First & Foremost is Inconsistent Database
- Consists of Missing Values and Blank Spaces in it.
- Classification becomes difficult
- Statistical Reports generated by considering various Systems approaches which are not concluded from same attributes
- Prediction to be done by considering all the attribute values which are indeed specific .
- Accuracy levels of predicted values obtained through Algorithms or generated systems are not rectified by any expert nor acknowledged by any authority.
- Considering many number of classes for prediction really misleads



Architecture of a typical expert system for a particular problem domain.

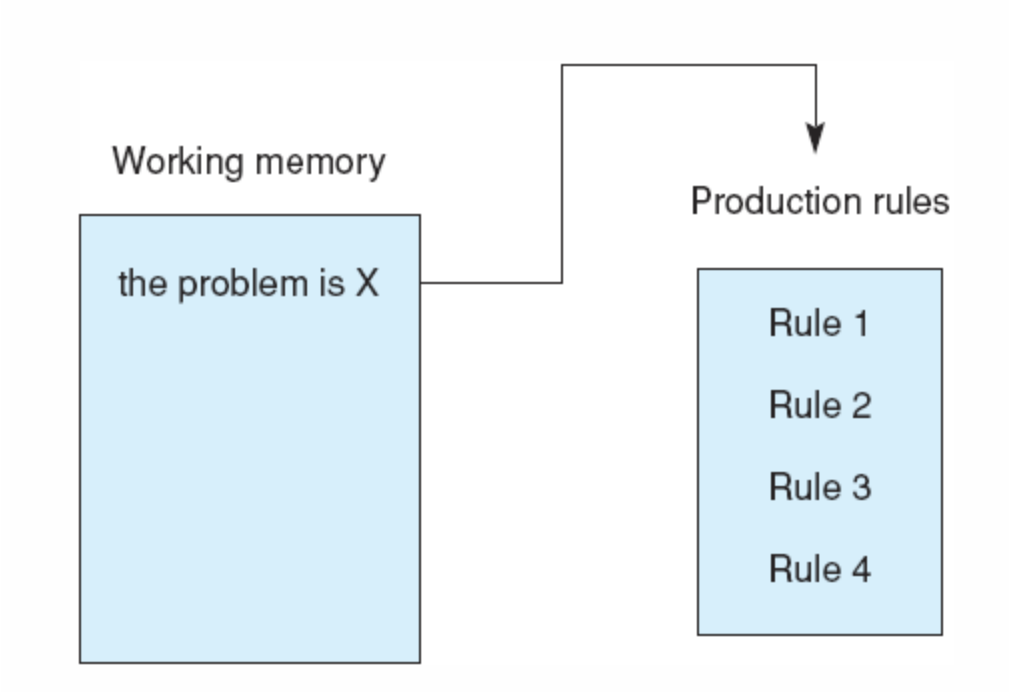
Rule 1: if
the engine is getting gas, and
the engine will turn over,
then
the problem is spark plugs.

Rule 2: if
the engine does not turn over, and
the lights do not come on
then
the problem is battery or cables.

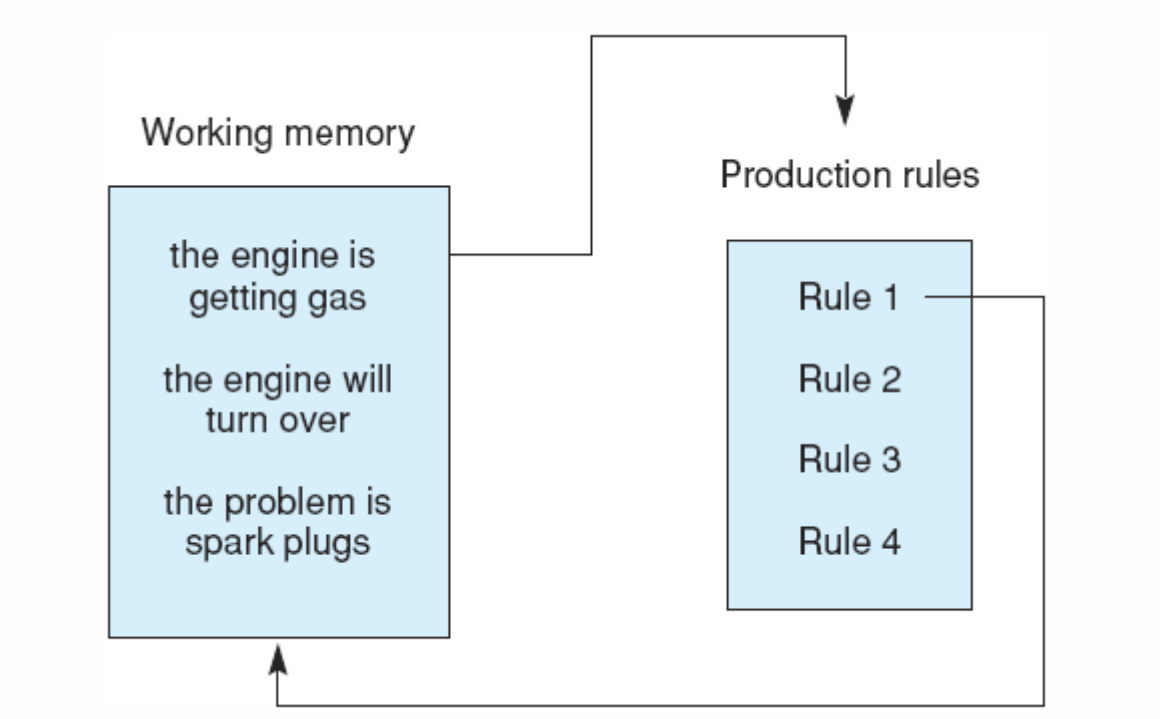
Rule 3: if
the engine does not turn over, and
the lights do come on
then
the problem is the starter motor.

Rule 4: if
there is gas in the fuel tank, and
there is gas in the carburetor
then
the engine is getting gas.

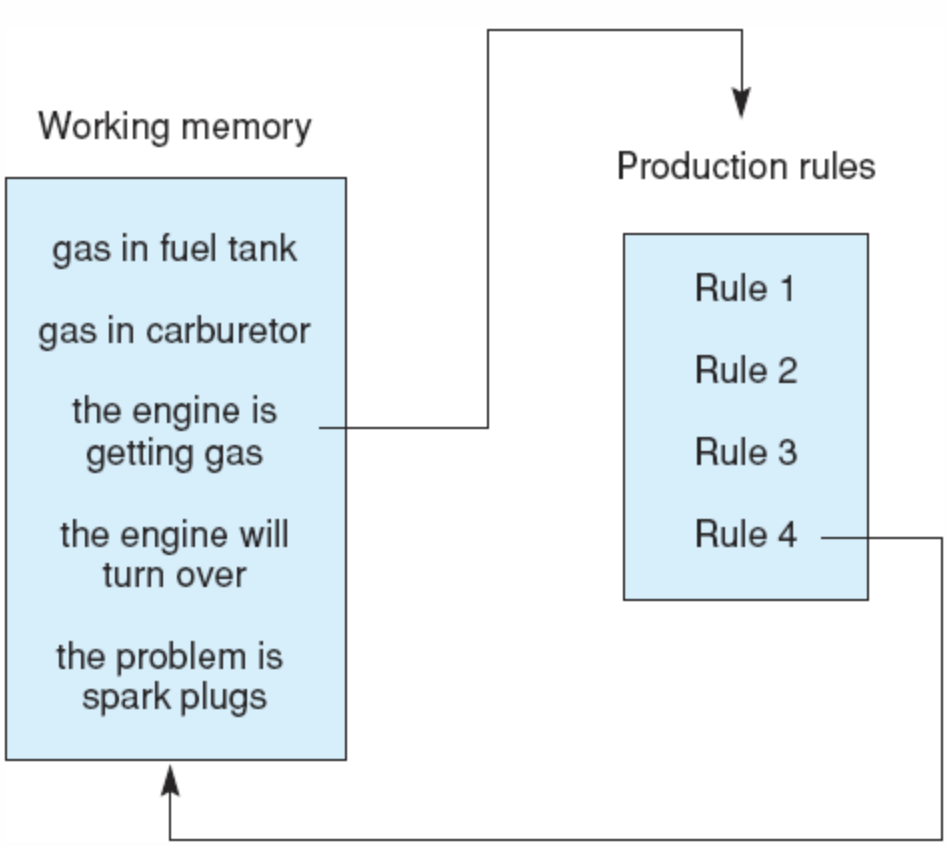
The production system at the start of a consultation in the car diagnostic example.



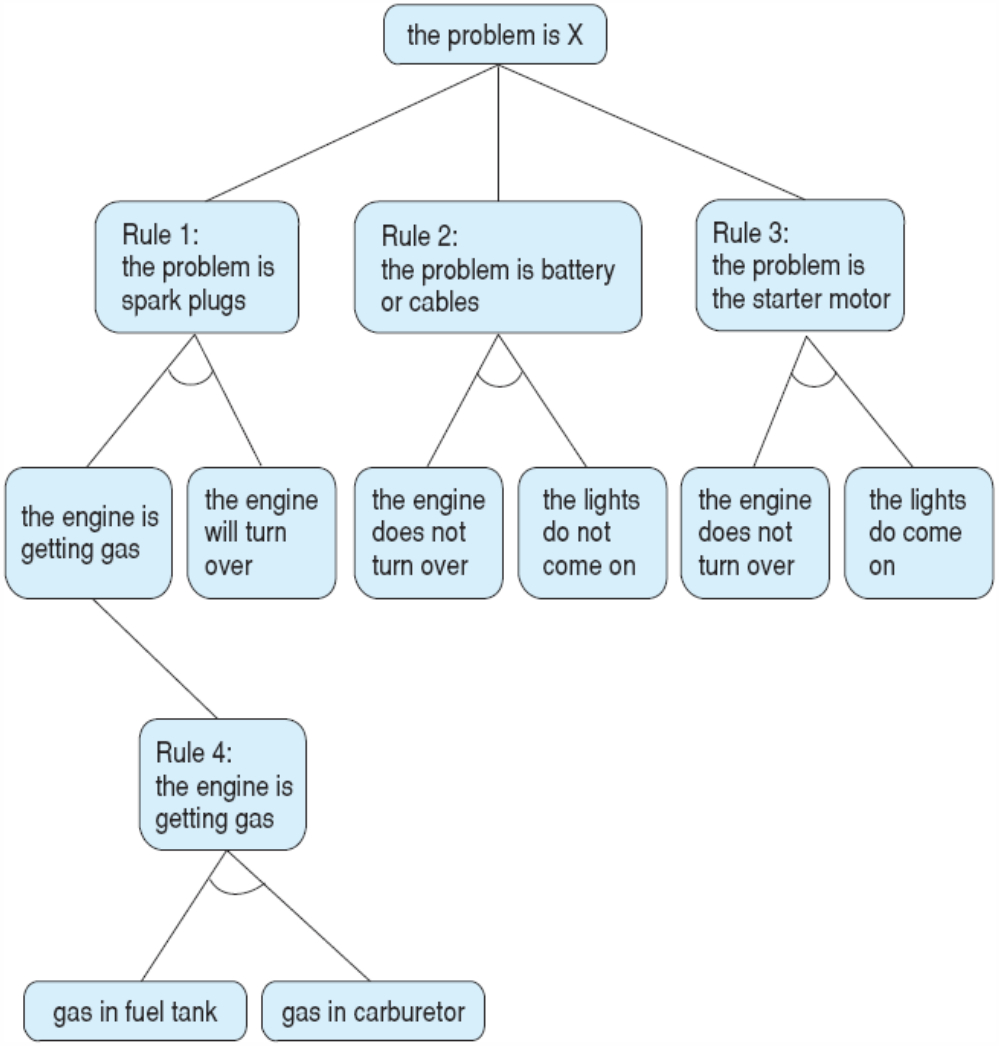
The production system after Rule 1 has fired.



The system after Rule 4 has fired. Note the stack-based approach to goal reduction.



The and/or graph searched in the car diagnosis example, with the conclusion of Rule 4 matching the first premise of Rule 1.



The following dialogue begins with the computer asking the user about the goals present in working memory.

Gas in fuel tank?
Yes
Gas in carburetor?
Yes
Engine will turn over?
Why

It has been established that:
1. The engine is getting gas,
2. The engine will turn over,
Then the problem is the spark plugs.

How the engine is getting gas

This follows from rule 4:
if
gas in fuel tank, and
gas in carburetor **then**
engine is getting gas.
gas in fuel tank was given by the user
gas in carburetor was given by the user

Disadvantages of Rule Based System

- Chance of learning is less
- New Knowledge generation is impossible
- Cannot handle complex situations and strategies
- Manual feed of Data with respect to experts choice
- Time Delays

Questions to Deep Learners in E-Class Room

Lets construct a simple dynamic dataset and classify it accordingly.

Machine Learning and AI

- Improve task performance through Observation & Teaching/Training.
- Acquire knowledge automatically for use in a task.
- Learning as a key component in intelligence.

Questions to Deep Learners
in
E-Class Room
Intelligent Will Answer in Seconds

How could a person stay 7 Days without sleep.

Is my question complete/ incomplete?

Kinds of Learning

- Rote Learning
- Learning from Instruction
- Learning by Analogy
- Learning from Observation and Discovery
- Learning from Examples

Inductive Learning

A concept of discovering new rules and instances through previous examples and models

How System Learn?

- Supervised

  These are task driven (Regression & Classification)

- Unsupervised

  These are Data driven (Clustering)

- Reinforcement

  Model reacts to the environment based on the input and output

Learning Model Algorithms

- Rule induction
  - E.g., Decision trees
- Knowledge based
  - E.g., Using a Domain Theory (Rough Sets)
- Statistical
  - E.g., Naïve Bayes, Nearest Neighbor, Support Vector Machines

Applications

- Language/speech
  - Machine Translation
  - Summarization
  - Grammars
- IR
  - Text Categorization, Relevance Feedback
- Medical
  - Assessment of Illness Severity
- Vision
  - Face Recognition, Digit Recognition, Outdoor Scene Recognition
- Security
  - Intrusion Detection, Network Traffic, Credit Fraud
- Social networks
  - Email Traffic

Language Tasks

- Text summarization
  - Task: given a document which sentences could serve as the summary
  - Training data: summary + document pairs
  - Output: rules which extract sentences given an unseen document
- Grammar induction
  - Task: produce a tree representing syntactic structure given a sentence
  - Training data: set of sentences annotated with parse tree
  - Output: rules which generate a parse tree given an unseen sentence

IR Task

- Text categorization
  - http://www.yahoo.com
  - Task: given a web page, is it news or not?
    - Binary classification (yes, no)
    - Classify as one of business & economy, news &media, computer
  - Training data: documents labeled with category
  - Output: a yes/no response for a new document; a category for a new document

Medical

- Task: Does a patient have heart disease (on a scale from 1 to 4)
- Training data:
  - Age, sex, cholesterol, chest pain location, chest pain type, resting blood pressure, smoker?, fasting blood sugar, etc.
  - Characterization of heart disease (0,1-4)
- Output:
  - Given a new patient, classification by disease

General Approach

- Formulate task prior model (parameters, structure)
- Obtain data and how to represent (attribute/value pairs)
- Annotate data, if needed
- Learn /Build /Refine model with data (training)
- Use model for classification or prediction on unseen data (testing)
- Measure Accuracy

Issues

- Representation
  - How to map from a representation in the domain to a representation used for learning?
- Training data
  - How can training data be acquired?
- Amount of training data
  - How well does the algorithm do as we vary the amount of data?
- Which attributes influence learning most?
- Does the learning algorithm provide insight into the generalizations made?

Classification Learning

- Input: a set of attributes and values
- Output: discrete valued function
  - Learning a continuous valued function is called regression
- Binary or boolean classification: category is either true or false

Learning Decision Trees

- Each node tests the value of an input attribute
- Branches from the node correspond to possible values of the attribute
- Leaf nodes supply the values to be returned if that leaf is reached

Example

- http://www.ics.uci.edu/~mlearn/MLSummary.html
- Iris Plant Database
- Which of 3 classes is a given Iris plant?
  - Iris Setosa
  - Iris Versicolour
  - Iris Virginica
- Attributes
  - Sepal length in cm
  - Sepal width in cm
  - Petal length in cm
  - Petal width in cm

Summary Statistics:

	Min	Max	Mean	SD	ClassCorrelation
sepal length:	4.3	7.9	5.84	0.83	0.7826
sepal width:	2.0	4.4	3.05	0.43	-0.4194
petal length:	1.0	6.9	3.76	1.76	0.9490 (high!)
petal width:	0.1	2.5	1.20	0.76	0.9565 (high!)

- Rules to learn
  - If sepal length > 6 and sepal width > 3.8 and petal length < 2.5 and petal width < 1.5 then class = Iris Setosa
  - If sepal length > 5 and sepal width > 3 and petal length >5.5 and petal width >2 then class = Iris Versicolour
  - If sepal length <5 and sepal width > 3 and petal length ≥ 2.5 and ≤ 5.5 and petal width ≥ 1.5 and ≤ 2 then class = Iris Virginica

Data

	S-length	S-width	P-length	P-width	Class
1	6.8	3	6.3	2.3	Versicolour
2	7	3.9	2.4	2.2	Setosa
3	2	3	2.6	1.7	Verginica
4	3	3.4	2.5	1.1	Verginica
5	5.5	3.6	6.8	2.4	Versicolour
6	7.7	4.1	1.2	1.4	Setosa
7	6.3	4.3	1.6	1.2	Setosa
8	1	3.7	2.8	2.2	Verginica
9	6	4.2	5.6	2.1	Versicolour

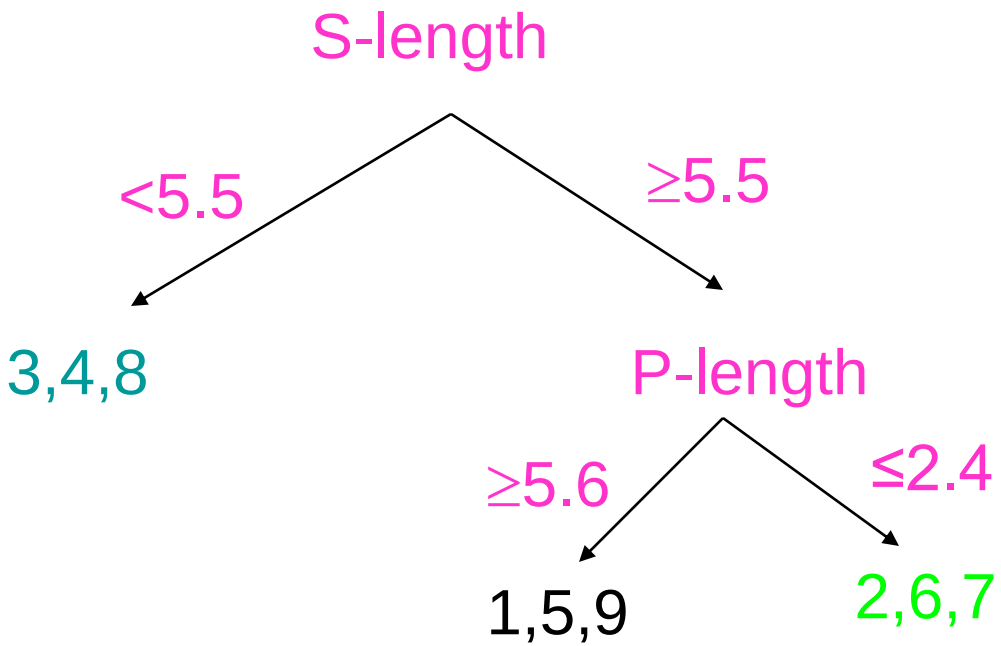
Data

	S-length	S-width	P-length		Class
1	6.8	3	6.3		Versicolour
2	7	3.9	2.4		Setosa
3	2	3	2.6		Verginica
4	3	3.4	2.5		Verginica
5	5.5	3.6	6.8		Versicolour
6	7.7	4.1	1.2		Setosa
7	6.3	4.3	1.6		Setosa
8	1	3.7	2.8		Verginica
9	6	4.2	5.6		Versicolour

Questions to Deep Learners in E-Class Room

- *Goal: Find the smallest decision tree consistent with the examples*
- Find the attribute that best splits examples
- Form tree with root = best attribute
- For each value v_i (or range) of best attribute
  - Selects those examples with best=v_i
  - Construct $subtree_i$ by recursively calling decision tree with subset of examples, all attributes except best
  - Add a branch to tree with label=v_i and subtree=$subtree_i$

Tree and Rules Learned



If S-length < 5.5., then Verginica

If S-length ≥ 5.5 and P-length ≥ 5.6, then Versicolour

If S-length ≥ 5.5 and P-length ≤ 2.4, then Setosa

Text Classification

- Is $text_i$ a finance new article?

Dow Closes Down 38, Nasdaq Gains 3

General Sense of Caution Pushes Stocks Mostly Lower; Dow Closes Down 38, Nasdaq Gains 3

The Associated Press

Overview | DJIA | S&P | NASDAQ

April 9, 2004

2 %
1 %
0 %
-1 %

10:00 noon 2:00 4:00

Find Quote

Scan Your Portfolio

April 8, 8:56 pm ET — Investors concerned about the deteriorating situation in Iraq looked past solid earnings from General Electric Co. and Yahoo! Inc. Thursday, sending stocks mostly lower and leaving Wall Street with a loss for the holiday-shortened week.

The Dow Jones industrial average was down 38.12, or 0.4 percent, at 10,442.03. The Dow had gained more than 60 points in early trading before falling back. Broader stock indicators were narrowly mixed. The Standard & Poor's 500 index was down 1.21, or 0.1 percent, at 1,139.32, while the Nasdaq composite index gained 2.64, or 0.1 percent, to 2,052.88.

While GE, which posted earnings in line with Wall Street estimates, is seen as a gauge of the overall economy due to the conglomerate's diverse businesses, the threat of increased violence in Iraq and possible consequences from terrorism kept investors from making large bets.

"The market is now in a duel between good economic numbers, good earnings, and the situation in Iraq," said Peter Cardillo, chief strategist and senior vice president at S.W. Bach & Co. "With the market trading at the upper end of its trading range for the year, it induces people to take some money off the table."

Trading was quiet and volume light, with many investors and traders taking time off for the holidays. The stock market was scheduled to close for Good Friday.

For the week, the Dow lost 0.3 percent, while the S&P 500 and Nasdaq both dropped

Positive

UConn defense knocks out Georgia Tech for second title

By Jack Carey, USA TODAY

SAN ANTONIO — Dominating inside, outside and especially defensively, Connecticut roared to its second men's basketball national championship in six years Monday night, rolling past Georgia Tech 82-73.

Negative

Jim Calhoun, center, and the rest of the UConn Huskies celebrate the school's second national title.

By Sue Ogrocki, AP

With the UConn women playing Tennessee for the national crown Tuesday night in New Orleans, Connecticut has a chance to become the first school to hold both basketball titles in the same year.

In its unexpected run to the final, Georgia Tech (28-10) had five games decided in the closing seconds, but Monday's contest was over early. (**Related item:** Box score)

With All-America center Emeka Okafor (24 points, 15 rebounds) virtually unstoppable down low and smooth guard Ben Gordon scoring 21, the Huskies (33-6) were too much for Tech, which couldn't find the range until late against UConn's aggressive defenders.

Dow Closes Down 38, Nasdaq Gains 3

General Sense of Caution Pushes Stocks Mostly Lower; Dow Closes Down 38, Nasdaq Gains 3

The Associated Press

April 8, 8:56 pm ET — Investors concerned about the deteriorating situation in Iraq looked past solid earnings from General Electric Co. and Yahoo! Inc. Thursday, sending stocks mostly lower and leaving Wall Street with a loss for the holiday-shortened week.

The Dow Jones industrial average was down 38.12, or 0.4 percent, at 10,442.03. The Dow had gained more than 60 points in early trading before falling back. Broader stock indicators were narrowly mixed. The Standard & Poor's 500 index was down 1.21, or 0.1 percent, at 1,139.32, while the Nasdaq composite index gained 2.64, or 0.1 percent, to 2,052.88.

While GE, which posted earnings in line with Wall Street estimates, is seen as a gauge of the overall economy due to the conglomerate's diverse businesses, the threat of increased violence in Iraq and possible consequences from terrorism kept investors from making large bets.

"The market is now in a duel between good economic numbers, good earnings, and the situation in Iraq," said Peter Cardillo, chief strategist and senior vice president at S.W. Bach & Co. "With the market trading at the upper end of its trading range for the year, it induces people to take some money off the table."

Trading was quiet and volume light, with many investors and traders taking time off for the holidays. The stock market was scheduled to close for Good Friday.

For the week, the Dow lost 0.3 percent, while the S&P 500 and Nasdaq both dropped 0.2 percent. The losses followed the market's best week of 2004

20 attributes

- Investors 2
- Dow 2
- Jones 2
- Industrial 1
- Average 3
- Percent 5
- Gain 6
- Trading 8
- Broader 5
- stock 5
- Indicators 6
- Standard 2
- Rolling 1
- Nasdaq 3
- Early 10
- Rest 12
- More 13
- first 11
- Same 12
- The 30

UConn defense knocks out Georgia Tech for second title

By Jack Carey, USA TODAY

SAN ANTONIO — Dominating inside, outside and especially defensively, Connecticut roared to its second men's basketball national championship in six years Monday night, rolling past Georgia Tech 82-73.

Jim Calhoun, center, and the rest of the UConn Huskies celebrate the school's second national title.

By Sue Ogrocki, AP

With the UConn women playing Tennessee for the national crown Tuesday night in New Orleans, Connecticut has a chance to become the first school to hold both basketball titles in the same year.

In its unexpected run to the final, Georgia Tech (28-10) had five games decided in the closing seconds, but Monday's contest was over early. (**Related item:** Box score)

With All-America center Emeka Okafor (24 points, 15 rebounds) virtually unstoppable down low and smooth guard Ben Gordon scoring 21, the Huskies (33-6) were too much for Tech, which couldn't find the range until late against UConn's aggressive defenders.

20 attributes

- Men's
- Basketball
- Championship
- UConn Huskies
- Georgia Tech
- Women
- Playing
- Crown
- Titles
- Games
- Rebounds
- All-America
- early
- rolling
- Celebrates
- Rest
- More
- First
- The
- same

Questions to Deep Learners in E-Class Room

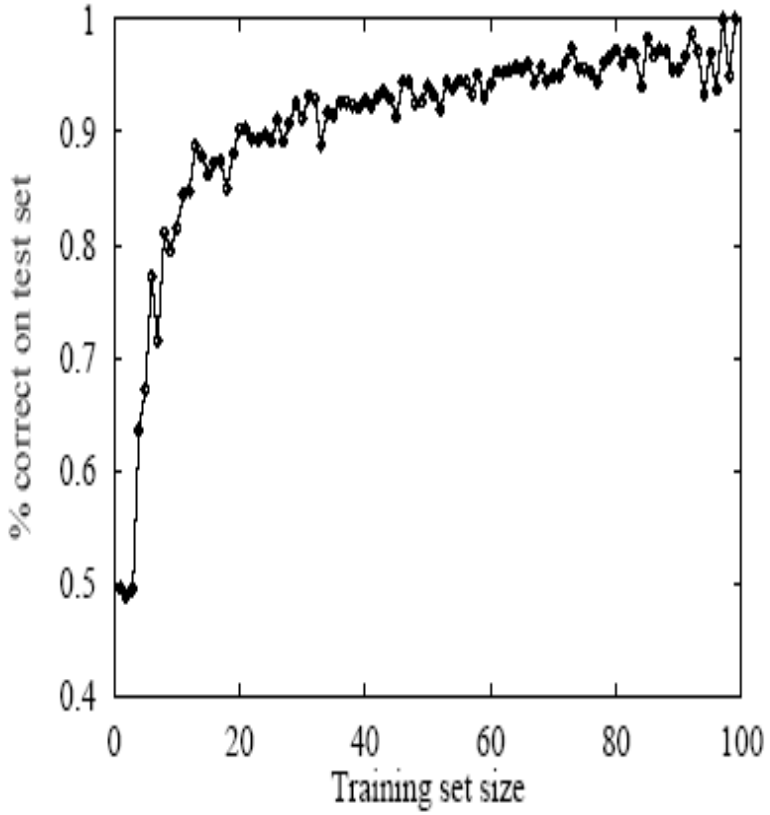
- What if class is discrete valued, not binary?
- What if an attribute has many values (e.g., 1 per instance)?

Training vs. Testing

A learning algorithm is good if it uses its learned hypothesis to make accurate predictions on unseen data

- Collect a large set of examples (with classifications)
- Divide into two disjoint sets: the training set and the test set
- Apply the learning algorithm to the training set, generating hypothesis h
- Measure the percentage of examples in the test set that are correctly classified by h
- Repeat for different sizes of training sets and different randomly selected training sets of each size.

Learning curve = % correct on test set as a function of training set size



Division into 3 sets

- Inadvertent peeking
  - Parameters that must be learned (e.g., how to split values)
  - Generate different hypotheses for different parameter values on training data
  - Choose values that perform best on testing data
  - Why do we need to do this for selecting best attributes?

Overfitting

- Learning algorithms may use irrelevant attributes to make decisions

K-fold Cross Validation

- To reduce overfitting
- Run k experiments
  - Use a different 1/k of data for testing each time
  - Average the results
- 5-fold, 10-fold, leave-one-out

Ensemble Learning

- Learn from a collection of hypotheses
- Majority voting
- Enlarges the hypothesis space

Questions to Deep Learners
in
E-Class Room

Lets work on regression/ prediction

Thank You