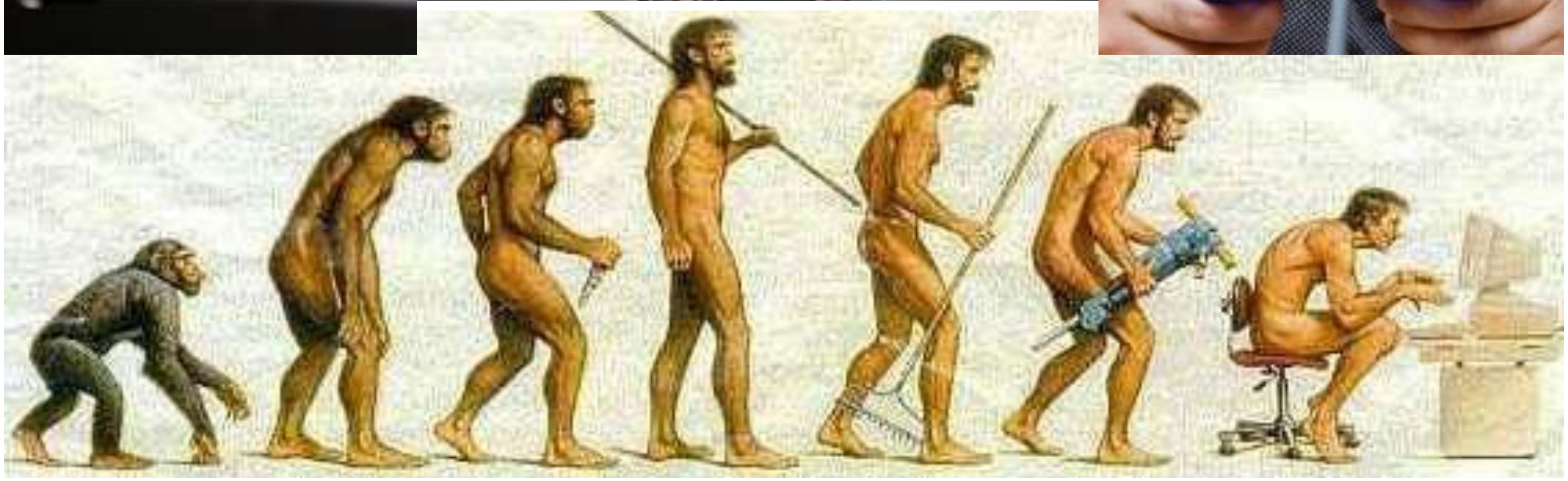
Machine learning
III B.Tech II Semester CSE JNTUH
CS601PC

J.Kamal Vijetha
Assistant Professor
KMIT

Machine Learning



Charles **Darwin's theory of evolution** states that **evolution** happens by natural selection. Individuals in a species show variation in physical characteristics. This variation is because of differences in their genes[2].

Machine Learning

- Learning ↔ Intelligence

  (Def: *Intelligence is the ability to learn and use concepts to solve problems.*)

- Machine Learning ↔ Artificial Intelligence
  - Def: *AI is the science of making machines do things that require intelligence if done by men* (Minsky 1986)
  - Def: *Machine Learning is an area of AI concerned with development of techniques which allow machines to learn*

- Why Machine Learning? ↔ Why Artificial Intelligence?

  ≡ To build machines exhibiting intelligent behaviour (i.e., able to reason, predict, and adapt) while helping humans work, study, and entertain themselves

Machine Learning

- Human Learning ↔ Machine Learning
  - human-logic inspired problem solvers (e.g., rule-based reasoning)
  - biologically inspired problem solvers (e.g., Neural Networks)
    - supervised learning - generates a function that maps inputs to desired outputs
    - unsupervised learning - models a set of inputs, labelled examples are not available
  - learning by education (e.g., reinforcement learning, case-based reasoning)
- General Problem Solvers vs. Purposeful Problem Solvers
  - emulating general-purpose human-like problem solving is impractical
  - restricting the problem domain results in 'rational' problem solving
  - example of General Problem Solver: Turing Test
  - examples of Purposeful Problem Solvers: speech recognisers, face recognisers, facial expression recognisers, data mining, games, etc.
- Application domains: security, medicine, education, finances, genetics, etc.

UNIT - I

Introduction - Well-posed learning problems, designing a learning system, Perspectives and issues in machine learning

Concept learning and the general to specific ordering – introduction, a concept learning task, concept learning as search, find-S: finding a maximally specific hypothesis, version spaces and the candidate elimination algorithm, remarks on version spaces and candidate elimination, inductive bias.

Decision Tree Learning – Introduction, decision tree representation, appropriate problems for decision tree learning, the basic decision tree learning algorithm, hypothesis space search in decision tree learning, inductive bias in decision tree learning, issues in decision tree learning.

UNIT 1: Introduction : Part A

UNIT - I

Introduction - Well-posed learning problems, designing a learning system, Perspectives and issues in machine learning

Concept learning and the general to specific ordering – introduction, a concept learning task, concept learning as search, find-S: finding a maximally specific hypothesis, version spaces and the candidate elimination algorithm, remarks on version spaces and candidate elimination, inductive bias.

1. ML Achievements at a glimpse

- Learn to recognise spoken words
- Predict recovery rates of pneumonia patients
- Detect fraudulent use of credit cards
- Drive autonomous vehicles on public highways
- Play games such as backgammon at levels approaching the performance of human world champions
- Classification of new astronomical structures

2. Various Disciplines influenced ML

- Artificial intelligence

  Learning symbolic representations of concepts. Machine learning as a search problem. Learning as an approach to improving problem solving. Using prior knowledge together with training data to guide learning.
- Bayesian methods

  Bayes' theorem as the basis for calculating probabilities of hypotheses. The naive Bayes classifier. Algorithms for estimating values of unobserved variables.
- Computational complexity theory

  Theoretical bounds on the inherent complexity of different learning tasks, measured in terms of the computational effort, number of training examples, number of mistakes, etc. required in order to learn.
- Control theory

  Procedures that learn to control processes in order to optimize predefined objectives and that learn to predict the next state of the process they are controlling.
- Information theory

  Measures of entropy and information content. Minimum description length approaches to learning. Optimal codes and their relationship to optimal training sequences for encoding a hypothesis.
- Philosophy

  Occam's razor, suggesting that the simplest hypothesis is the best. Analysis of the justification for generalizing beyond observed data.
- Psychology and neurobiology

  The power law of practice, which states that over a very broad range of learning problems, people's response time improves with practice according to a power law. Neurobiological studies motivating artificial neural network models of learning.
- Statistics

  Characterization of errors (e.g., bias and variance) that occur when estimating the accuracy of a hypothesis based on a limited sample of data. Confidence intervals, statistical tests.

❖ Various Disciplines influenced ML

- Machine Learning ↔ Artificial Intelligence
- Machine Learning ← Biology (e.g., Neural Networks, Genetic Algorithms)
- Machine Learning ← Cognitive Sciences (e.g., Case-based Reasoning)
- Machine Learning ← Statistics (e.g., Support Vector Machines)
- Macine Learning ← Probability Theory (e.g., Bayesian Networks)
- Machine Learning ← Logic (e.g., Inductive Logic Programming)
- Machine Learning ← Information Theory (e.g., used by Decision Trees)

3.Well-posed Learning Problems

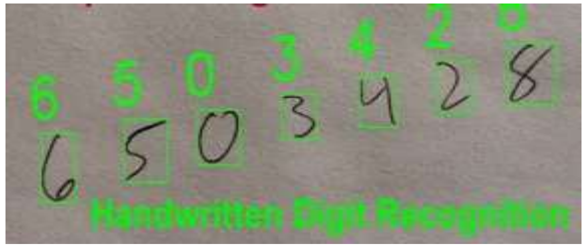
*A computer program is said to learn from experience **E** with respect to some class of **tasks T** and **performance measure P**, if its performance at tasks in T, as measured by P, improves by **experience E**.*

Checkers Learning problem:

T : checkers (draughts)

P : percent of games won in world tournament

E : Playing practice Games against Itself



Handwriting Recognition Learning Problem:

T : recognizing and classifying handwritten words with images

P : percent of words correctly classified

E: a database of handwritten words with given classifications

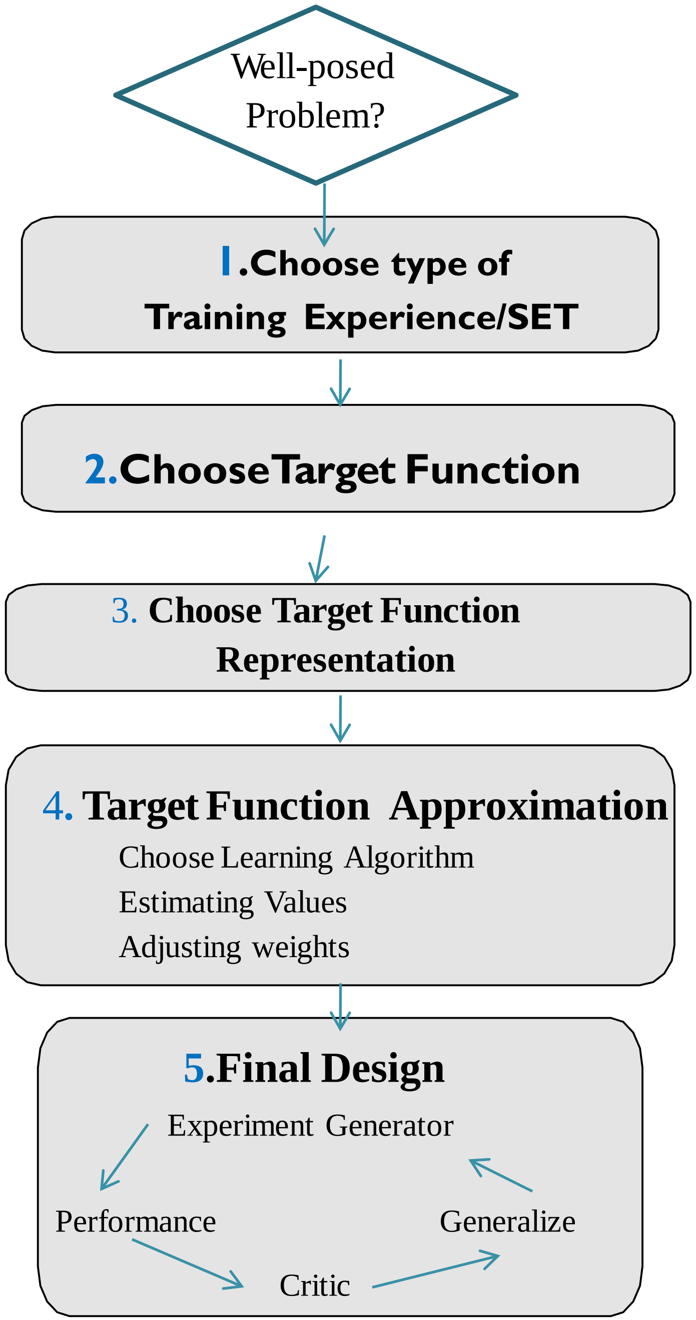
Robot driving Learning Problem:

T : driving on public four lane highways using vision sensors

P : average distance traveled before an error (as judged by human overseer)

E: a sequence of images and steering commands recorded while observing a human driver

4.Designing a Machine Learning System



For a well posed ML problem ,

based on the design issues and

approaches the different steps are used to measure the performance of the problem.

To complete the Design of learning Problem ,choose basic steps are

- Exact type of knowledge to be learned
- A representation for this target Knowledge
- A learning Mechanism

Step 1: **Choose type of Training Experience/dataSET**

The training set choosing is when we choose for an application .It impacts on the success and Failure of the system.

Training experience can be

Direct training – teacher-------------→BEST

Indirect training- no teacher

* Credit assignment --- disadvantage

Random Training - expert teacher

Step 2 ChooseTarget Function

- **Target Function V**
  What type of knowledge will be learned and how this will be used by the performance program
  (e.g., **choosing the best next move in checkers,** identifying people, C lassifying facial expressions into emotion categories)

Choose Move :

$$V: Board \rightarrow M$$

where B – Set of legal board states
M – Best move

$V: Board \rightarrow [-1, 1]$ is a general target function of a binary Classifier

- **Alternate Target Function V** used is

$$V: Board \rightarrow \mathbb{R}$$

ie., R is real Numbers change with mean median or mode

Step 3 Choose Target Function Representation

- Ideal Target Function is usually not known; machine learning algorithms learn an approximation of V, say V'
  - **be as close an approximation of V as possible**
  - require (reasonably) small amount of training data to be learned

$$\hat{V} \text{ or } V' : \text{Approximation } (V)$$

- Approximated Target Function V' is for a collection of rule is defined with a quadratic Polynomial or ANN
- If b is arbitrary board state in B then V(b) is

1. if b is a final board state that is won, then $V(b) = 100$
2. if b is a final board state that is lost, then $V(b) = -100$
3. if b is a final board state that is drawn, then $V(b) = 0$
4. if b is a not a final state in the game, then $V(b) = V(b')$, where b' is the best final board state that can be achieved starting from b and playing optimally until the end of the game (assuming the opponent plays optimally, as well).

$$V'(b) = w_0 + w_1x_1 + \ldots + w_nx_n$$

where $\langle x_1 \ldots\ldots x_n \rangle \equiv b \in$ B leagal *Board state* is an input state ie training set .This reduces the problem to learning (the most optimal) weights w.

Step 4 Target Function Approximation

Choose Learning Algorithm
Estimating Values
Adjusting weights

- To make Target Function V' learn

➢First need Set of training examples

$\langle b, V_{train}(b)\rangle$ → ordered pair

$\langle(x_1=3, x_2=0, x_3=1, x_4=0, x_5=0, x_6=0), +100\rangle$

x1: the number of the black pieces on the board
x2: the number of the red pieces on the board
x3: the number of the black major on the board
x4: the number of the red major on the board
x5: the number of black pieces threatened by red
x6: the number of red pieces threatened by black

Changing $x_2=0$ gives a fit for direct training
If is indirect training change the weights

➢ $V_{train}(b) \leftarrow V'(successor(b))$

if V´ tends to be more accurate for board positions closer to game's end

➢Learning Algorithm for choosing weights w_i to best fit the set of training examples

$\{\langle b, V_{train}(b)\rangle\} \equiv \{\langle b, V'(Successor(b))\rangle\}$

Best fit could be defined as minimizes the squared error E

$$E \equiv \sum_{\langle b, V_{train}(b)\rangle \in training\text{-}examples} (V_{train}(b) - V'(b))^2$$

$(V_{train}(b) - \hat{V}(b)) = 0$ -> No need to change weights
$(V_{train}(b) - \hat{V}(b)) > 0$ -> Increase weight values
$(V_{train}(b) - \hat{V}(b)) < 0$ -> Decrease weight values

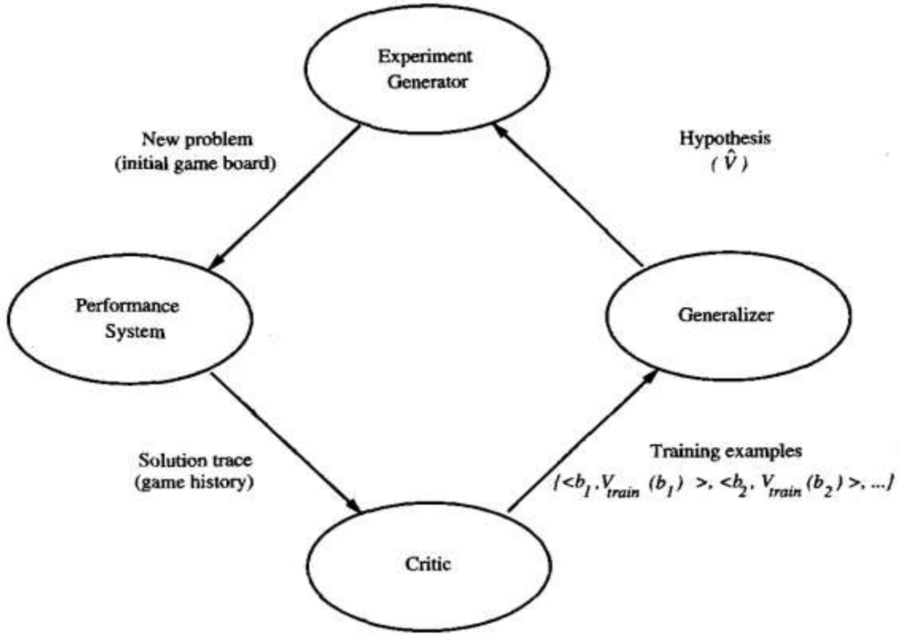
Least Mean Square:

For each $\langle b, V_{train}(b)\rangle$ use current weights to calculate V´(b). For each weight

$$w_i \leftarrow w_i + \eta(V_{train}(b) - V'(b))x_i$$

Where η is a small constant .01 that moderates the size of the weight update

Step 5 Final Design



6.Design of learning system example

Partial design of a checkers learning program:

- Task T: playing checkers
- Performance measure P: percent of games won in the world tournament
- Training experience E: games played against itself
- *Target function:* $V: Board \rightarrow \Re$
- *Target function representation*

$$\hat{V}(b) = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + w_6x_6$$

the target function value $V_{train}(b)$ is therefore $+100$.

$$\langle\langle x_1 = 3, x_2 = 0, x_3 = 1, x_4 = 0, x_5 = 0, x_6 = 0\rangle, +100\rangle$$

Rule for estimating training values.

$$V_{train}(b) \leftarrow \hat{V}(Successor(b))$$

LMS weight update rule.

For each training example $\langle b, V_{train}(b)\rangle$

- Use the current weights to calculate $\hat{V}(b)$
- For each weight w_i, update it as

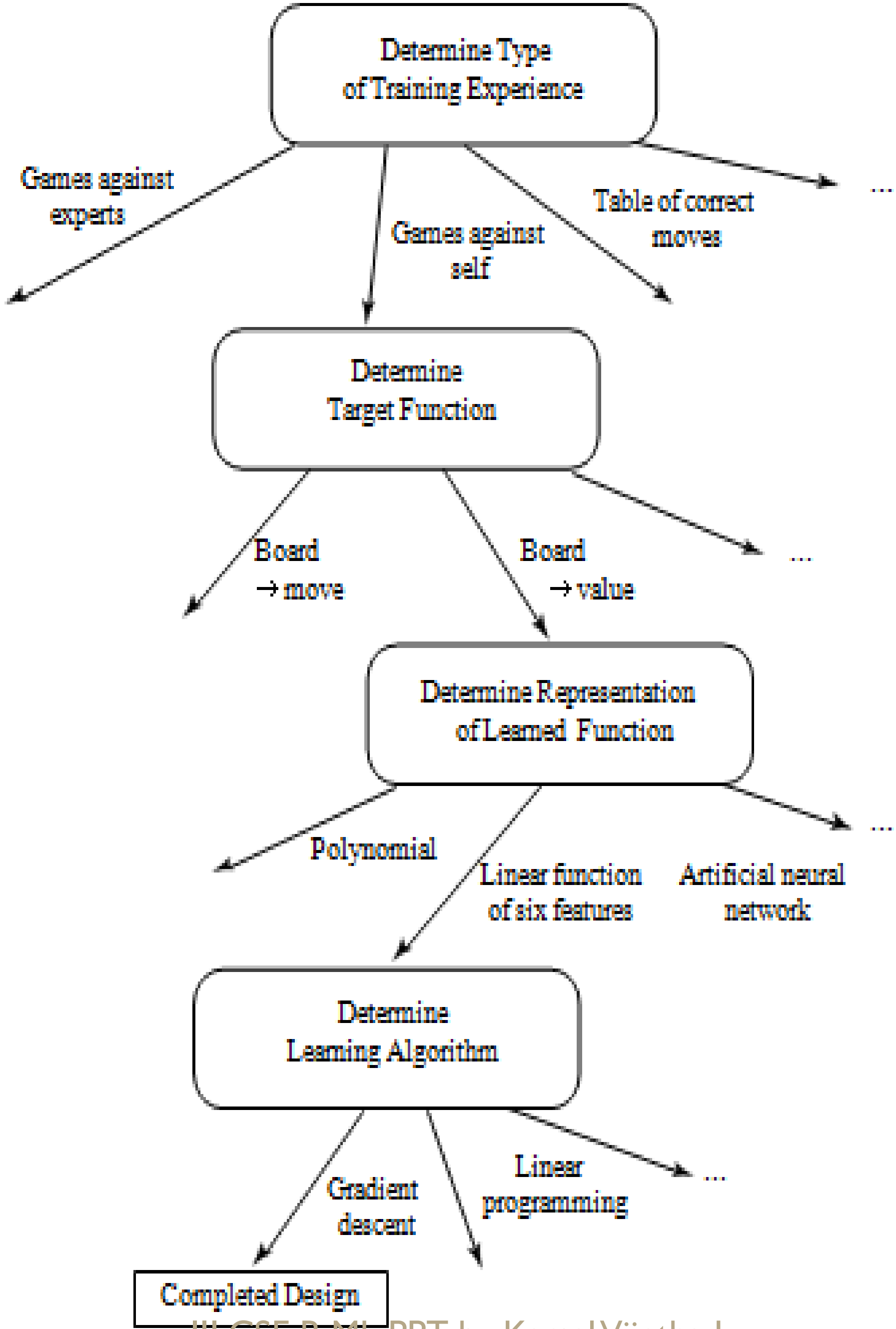
$$w_i \leftarrow w_i + \eta\ (V_{train}(b) - \hat{V}(b))\ x_i$$

Accuracy of win is maximum for the BEST FIT ie is for the values of v'(b) is low

Design choices made for Checkers Program of learning system example

Assumptions:

1. Single linear evaluation function
2. 6 specific board Features



7a. Perspective in ML

It involves searching a **very large space of possible hypotheses (best evidence for a problem) to determine one that best fits with** the observed data and any prior knowledge held by the learner.

For example:

- hypothesis space consists of all evaluation functions that can be represented by some choice of values for the weights w_0 through w_6.
- The LMS algorithm for fitting weights achieves this goal by iteratively tuning the weights, adding a correction to each weight
- for searching a hypothesis space defined by some underlying representation (e.g., linear functions, logical descriptions, decision trees, artificial neural networks).

The perspective of learning as a search problem in order to characterize learning methods by their search strategies and by the underlying structure of the search spaces.

7b. Issues in ML

- **What algorithms exist for learning general target functions from specific training examples?**
- In what settings will particular algorithms converge to the desired function, given sufficient training data?
- Which algorithms perform best for which types of problems and representations?
- **How much training data is sufficient?**
- What general bounds can be found to relate the confidence in learned hypotheses to the amount of training experience and the character of the learner's hypothesis space?
- **When and how can prior knowledge held by the learner guide the process of generalizing from examples?**
- Can prior knowledge be helpful even when it is only approximately correct?
- **What is the best strategy for choosing a useful next training experience, and how does the choice of this strategy alter the complexity of the learning problem?**
- **What is the best way to reduce the learning task to one or more function approximation problems?**
- Put another way, what specific functions should the system attempt to learn?
- Can this process itself be automated?
- **How can the learner automatically alter its representation to improve its ability to represent and learn the target function?**