

Introduction to Machine Learning

AICTE-ISTE Induction/Refresher Program sponsored
One Week FDP Phase-II on Machine Learning for all
organized by Department of ECE & CRI,
K.S.R.M. college of Engineering, Kadapa, Andhra Pradesh
(5 - 10 April 2021)

Mukesh A. Zaveri
Computer Engineering Department
Sardar Vallabhbhai National Institute of Technology, Surat
mazaveri@coed.svnit.ac.in



Artificial Intelligence

- To Create **Expert Systems**
 - ▶ The systems which **exhibit intelligent behavior, learn, demonstrate, explain, and advice** its users
- To **Implement Human Intelligence** in Machines
 - ▶ Creating **systems that understand, think, learn, and behave like humans**
- What Contributes to AI?
- AI is a science and technology based on **disciplines** such as
- Computer Science, Biology, Psychology, Linguistics, Mathematics, and Engineering
- A major thrust of AI is in the development of computer functions associated with **human intelligence, such as reasoning, learning, and problem solving**



Artificial Intelligence based Applications of next Era Google

source

- AI could help create hit songs?
 - ▶ AI searched through millions of conversations, newspaper titles, and lectures based on the main theme of the song: heartbreak.
- IBM's Chef Watson
 - ▶ gives a glimpse of the creative potential for AI
 - ▶ Chef Watson's recipes might suggest **ingredient combinations and styles of dishes** humans would never have considered,
 - ▶ its ability to **analyse data and overlay scientific info** makes working in the kitchen alongside human chefs effective.
 - ▶ Chef Watson helps to **create a recipe to pick any ingredient** you wish to use in your culinary adventure



AI based Applications of next Era Google source

- Drive without a driver
 - ▶ guided by AI technologies and automatic learning?
 - ▶ Tesla was one of the first automotive brands to launch a self-driving vehicle, and Audi, Cadillac, and Volvo are already developing their own models.
- The Fortuneteller That Will Know It Before You
 - ▶ AI can assertively predict if someone is gay or straight based on photographs of their face.
 - ▶ A Stanford University study
 - ▶ IBM has a solution that uses predictive analytics to identify personnel retention problems.
- A Superhuman Doctor
 - ▶ AI and deep learning, doctors can promptly diagnose cancer, before its too late.
- The Smartest Investor



AI based Applications of next Era Google source

- Photographs That Become Purchases

- ▶ Amazon incorporated a **visual search option** on its mobile application
- ▶ take a photo of the item you want, and it will show you something very similar or identical
- ▶ buy it right away and its already in your hands.

- A Better World

- ▶ AI can help us to **prevent future damage** and better **understand how to address developmental needs** while focusing on sustainability.
- ▶ Microsoft using AI to study **land-use patterns** with terrain maps.
- ▶ A deep **understanding of these patterns** allows it to make better decisions on the use of the land and implement proper preservation techniques.
- ▶ Scientists would be able to use the information obtained to preserve biodiversity and the ecosystem.
- ▶ EarthCube is one of these projects.



AI based Applications of next Era Google source

- Life On Other Planets?

- ▶ NASA is using AI to look for life on other planets,
- ▶ Mars 2020: the mission where the Red Planet will be explored more thoroughly.

- The Marketing Guru

- ▶ AI is becoming the right hand of marketers and retailers around the world.
- ▶ Through AI and machine learning, the most **profitable audiences** can be found for any ad.
- ▶ Intelligent algorithms are able to **learn, detect, and predict** which types of users are more likely to become clients at a lower cost per acquisition.

- The Sales And Customer Service Genie

- ▶ AI is also becoming a retailer's best friend
- ▶ AI platforms that **monitor the habits** of sales representatives, giving them **recommendations** to improve their performance.



AI based Applications of next Era Google source

- A Non-Human Language

- ▶ Facebook is using machine learning to teach its chatbots to converse and negotiate amongst themselves (chatbot to chatbot).

- The Best Flight

- ▶ Through predictive analytics driven by AI,
- ▶ the app is able to predict price patterns and alert travelers of the cheapest times to buy flights to their destinations.

- In Charge of Your Accounting And Financing

- ▶ automate your accounting system and financial reports?
- ▶ upload your receipts, and the platform will convert them into a legible format for computers, encrypt them, and then put them into your account.
- ▶ It will also learn to monitor invoices, sales, and costs, as well as your liquidity.



AI in Robotics

- **Vision:** AI is helping robots **detect items** they have never seen before and **recognize objects** with far greater detail
- **Grasping:** robots are also grasping items they've never seen before with AI and machine learning helping them determine the **best position and orientation to grasp an object**
- **Motion Control:** machine learning helps robots with **dynamic interaction and obstacle avoidance** to maintain productivity
- **Data:** AI and machine learning both help robots **understand physical and logistical data patterns** to be proactive and act accordingly



AI in Robotics

- Japanese Hotel Run Almost Entirely By Robots
- Robots replace waiters in Chennai Restaurant
- Room service is delivered by robots at hotels in Singapore
- robots: **working faster and more reliably** than their human counterparts
- performing tasks beyond human capability altogether, e.g. **microscopically precise assembly**



Soft Computing - Real Life Example

- Disclaimer - with views presented

Birth / Death Certificate

HSC passing

Licenses at RTO

Ruling in Judiciary

Public Distribution (Ration shop)

City Survey office

Water (...) connection

Hospital / Doctor

Corporation / Government offices

School Admission

UG / PG Admission

Income Tax / VAT Refund

FIR at Police Station

Loan passing - Bank

Land transfer

Plan approval

Import / Export Licenses

Increment at Service



Soft Computing: Tool and Algorithm

- Tools:
 - ▶ Money
 - ▶ Contact
- Algorithm:
 - ▶ Not Known;
 - ▶ Human psychology
- Two sides of coin
- Soft Computing for



Soft Computing

- What is Soft Computing?
- Hard Computing Vs. Soft Computing
- Soft Computing Methods
 - ▶ Neural Network
 - ▶ Genetic Algorithm
 - ▶ Fuzzy Logic
- How to integrate / use for specific Application?
- Any other technique to be used with Soft Computing technique?



Soft Computing Techniques

- How to select soft computing technique?
- Neural Networks
 - ▶ learning, classification, optimization
- Fuzzy Logic
 - ▶ forming imprecision and reasoning on semantic or linguistic level
- Genetic Algorithm
 - ▶ exploring set of all possible solutions
- Probabilistic Reasoning
 - ▶ deals with uncertainty
- substantial areas of overlapping among different techniques
- they are complementary rather than competitive



Why Optimization?

- To achieve the goal with
 - ▶ Minimum computations
 - ▶ Minimum search
 - ▶ Efficient and Fast enough (for real time application)
 - ▶ At par with ideal goal (optimal solution)
- Optimization Techniques
 - ▶ Neural Network
 - ▶ Genetic Algorithm



Need for Learning based Application

- Applications
- Banking and Financial Services
 - ▶ macro economic conditions
 - ▶ changing market dynamics
 - ▶ product centric to customer focused
 - ▶ data driven transformation
- Insurance
 - ▶ emerging technologies, including drones,
 - ▶ Big Data and Analytics to transform
 - ▶ claims processing,
 - ▶ enhance risk management and
 - ▶ streamline overall operations



Need for Learning based Application

- Media & Entertainment
 - ▶ consumer behavior
 - ▶ wide variety of media and entertainment solutions
- Process Manufacturing
 - ▶ Industrial Internet of Things (IIoT),
 - ▶ Smart Manufacturing, predictive analytics
 - ▶ optimize supply chain operations,
 - ▶ drive down operating costs, and increase plant yields
- Engineering & Construction
 - ▶ integrate operational parameters with enterprise-level decision-making processes
 - ▶ establish more efficient, integrated workflows,
 - ▶ risk management practices to stay profitable in the face of volatility



Machine Learning

- How does a machine learn?
 - ▶ Build models and make inference from a sample
 - ▶ Model having parameters and learning algorithms optimizing these parameters using past data (samples)
 - ▶ Learning rule - Knowledge extraction
- Data Mining: Learning association - association rule $P(Y|X)$
- Distinction among customers, web personalization
- Classification: credit scoring - risk calculation and classify between low risk and high risk
- Pattern recognition
 - ▶ Character recognition, word, sentences and syntax, semantics
 - ▶ Speech age, gender, accent, pronounce
 - ▶ Lip movement recording with speech recognition needs sensor fusion



Machine Learning Techniques

- Decision tree learning
- Association rule learning
- Artificial neural networks - Deep learning
- Inductive logic programming
- Support vector machines
- Clustering
- Bayesian networks
- Reinforcement learning
- Representation learning
- Similarity and metric learning
- Sparse dictionary learning
- Genetic algorithms
- Rule-based machine learning - Learning classifier systems



Machine Learning

- Regression
 - ▶ Y price of car and X affecting car's attributes $y = wx + w_0$
- Classification
 - ▶ Input X and Output Y needs to learn the mapping $y = g(x|\theta)$
- Regression and Classification - Supervised learning
- Regression Y is number and in classification Y is class code
- Regression g is regression function and in classification g is discriminant function
- Unsupervised Learning
 - ▶ Only input data
 - ▶ needs density estimation
 - ▶ called clustering



Bayes decision theory

- Decision under uncertainty
- Data comes from unknown process so process can be modelled as random process
- Process may be deterministic but no complete knowledge so model as random and probability theory can be used to analyze it
- Calculate the probabilities of the classes
- Minimize expected risk
- Bayesian network to represent dependencies among random variables
- Mathematical framework of making inference from data



- Bank application: Credit scoring
- Observable: x_1 and x_2 income and saving
- Non observable: State of economy, full detail about customer, his moral codes, intention etc.
- Outcome: $C = 1$ high risk and $C = 0$ low risk
- for any new data $X_1 = x_1$ and $X_2 = x_2$ knowing $P(C|X_1, X_2)$ we can choose
 - ▶ $C = 1$ if $P(C = 1|x_1, x_2) > 0.5$ and $C = 0$ otherwise
 - ▶ $C = 1$ if $P(C = 1|x_1, x_2) > P(C = 0|x_1, x_2)$ and $C = 0$ otherwise
- Probability of error is $1 - \max(P(C = 1|x_1, x_2), P(C = 0|x_1, x_2))$



Classification using Bayes theory

- Let, observed variables $\mathbf{x} = [x_1, x_2]^T$
- Need to calculate $P(C|\mathbf{x})$
- Bayes's rule $P(C|\mathbf{x}) = \frac{P(C)p(\mathbf{x}|C)}{p(\mathbf{x})}$
- $P(C = 1)$ is prior probability (regardless of \mathbf{x} value, before looking at observable \mathbf{x}) $P(C = 0) + P(C = 1) = 1$
- $p(\mathbf{x}|C)$ is called class likelihood
- It is conditional probability event belonging to C has the associated observation value \mathbf{x} . It is what the data tells us regarding the class.
- $p(\mathbf{x})$ is evidence.
- The marginal probability that an observation \mathbf{x} is seen regardless of whether it is a positive or negative example

$$p(\mathbf{x}) = p(\mathbf{x}|C = 1)P(C = 1) + p(\mathbf{x}|C = 0)P(C = 0)$$



Classification using Bayes theory

- Calculate posterior probability $P(C|x)$ after seen observation
- How to estimate $P(C)$ and $p(x|C)$
- In general, C_i $i = 1, \dots, K$ with $P(C_i) \geq 0$ and $\sum_{i=1}^K P(C_i) = 1$
- $p(x|C_i)$ is the probability of seeing x as the input when it is known to belong to class C_i
- The posterior probability of class C_i

$$P(C_i|x) = \frac{p(x|C_i)P(C_i)}{p(x)} = \frac{p(x|C_i)P(C_i)}{\sum_{k=1}^K p(x|C_k)P(C_k)}$$

- For minimum error Bayes' classifier chooses the class with highest posterior probability choose C_i if $P(C_i|x) = \max_k P(C_k|x)$



Losses and Risk

- An accepted low risk customer increases profit
- A rejected high risk customer decreases loss
- The loss for a high risk customer erroneously accepted different from the gain for an erroneously rejected low risk customer
- Let, α_i action as decision to assign the input to class C_i (K actions $\alpha_i, i = 1, \dots, k$)
- λ_{ik} as the loss incurred for taking action α_i when the input actually belongs to C_k
- The expected risk for taking action α_i is

$$R(\alpha_i|x) = \sum_{k=1}^K \lambda_{ik} P(C_k|x)$$

- Choose the action with minimum risk i.e. choose α_i if $R(\alpha_i|x) = \min_k R(\alpha_k|x)$



Losses and Risk

- Special case zero-one loss

$$\lambda_{ik} = \begin{cases} 0 & \text{if } i = k \\ 1 & \text{if } i \neq k \end{cases}$$

$$R(\alpha_i|x) = \sum_{k \neq i} \lambda_{ik} P(C_k|x) = 1 - P(C_i|x)$$

- For applications misclassification may have very high cost
- Additional action reject or doubt α_{k+1} is defined
- A possible loss function

$$\lambda_{ik} = \begin{cases} 0 & \text{if } i = k \\ \lambda & \text{if } i = k + 1 \\ 1 & \text{otherwise} \end{cases}$$



Parametric Methods for Classification and Regression

- Parametric approach for classification and regression
- Samples drawn from some distribution that obeys a known model
- Small number of parameters e.g. mean, variance, sufficient statistics
- Estimate these parameters and obtain estimated distribution to make a decision called maximum likelihood estimation (MLE)
- Independent and identically distributed (iid) sample $\mathcal{X} = \{x^t\}_{t=1}^N$
- x^t drawn from $p(x|\theta) \rightarrow x^t \sim p(x|\theta)$

$$l(\theta) \equiv p(\mathcal{X}|\theta) = \prod_{t=1}^N p(x^t|\theta)$$

finding θ that makes \mathcal{X} the most likely to be drawn log likelihood

$$\mathcal{L}(\theta|\mathcal{X}) \equiv \log l(\theta|\mathcal{X}) = \sum_{t=1}^N \log p(x^t|\theta)$$



Parametric Method: Example

- Gaussian Density $x^t \sim \mathcal{N}(\mu, \sigma^2)$

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right]$$

$$\mathcal{L}(\mu, \sigma | \mathcal{X}) = -\frac{N}{2} \log(2\pi) - N \log \sigma - \frac{\sum_t (x^t - \mu)^2}{2\sigma^2}$$

- MLE for μ and σ are

$$\hat{\mu} = \frac{\sum_t x^t}{N} \quad \sigma^2 = \frac{\sum_t (x^t - \hat{\mu})^2}{N}$$



Parametric classification

- Using Bayes' rule

$$P(C_i|x) = \frac{p(x|C_i)P(C_i)}{p(x)} = \frac{p(x|C_i)P(C_i)}{\sum_{k=1}^K p(x|C_k)P(C_k)}$$

- Use the discriminant function

$$g_i(x) = p(x|C_i)P(C_i) \quad \text{or} \quad g_i(x) = \log p(x|C_i) + \log P(C_i)$$

$$p(x|C_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left[-\frac{(x - \mu_i)^2}{2\sigma_i^2} \right]$$

$$g_i(x) = -\frac{1}{2} \log 2\pi - \log \sigma_i - \frac{(x - \mu_i)^2}{2\sigma_i^2} + \log P(C_i)$$

- If priors are equal and variances are equal then $g_i(x) = -(x - \hat{\mu}_i)^2$
- Assign x to the class with the nearest mean: choose C_i if $|x - \hat{\mu}_i| = \min_k |x - \hat{\mu}_k|$



Parametric Regression

- $r = f(x) + \epsilon$
- numeric output r is sum of a deterministic function $f(x)$ of input and random noise
- $f(x)$ unknown and approximate by estimator $g(x|\theta)$
- Say, $\epsilon \sim \mathcal{N}(0, \sigma^2)$ then $p(r|x) \sim \mathcal{N}(g(x|\theta), \sigma^2)$
- Use MLE to learn parameters θ
- (x^t, r^t) are drawn from an unknown joint probability density
 $p(x, r) = p(r|x)p(x)$
- Sample $\mathcal{X} = \{x^t, r^t\}_{t=1}^N$

$$\mathcal{L}(\theta|\mathcal{X}) = \log \prod_{t=1}^N p(x^t, r^t) = \log \prod_{t=1}^N p(r^t|x^t) + \log \prod_{t=1}^N p(x^t)$$



Parametric Regression

- Ignoring second term, as estimator does not depend on it

$$\begin{aligned}\mathcal{L}(\theta|\mathcal{X}) &= \log \prod_{t=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{[r^t - g(x^t|\theta)]^2}{2\sigma^2} \right] \\ &= \log \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^N \exp \left[-\frac{1}{2\sigma^2} \sum_{t=1}^N [r^t - g(x^t|\theta)]^2 \right] \\ &= -N \log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{t=1}^N [r^t - g(x^t|\theta)]^2\end{aligned}$$

- Maximizing above is equivalent to minimizing

$$\frac{1}{2} \sum_{t=1}^N [r^t - g(x^t|\theta)]^2$$

- θ that minimizes is called least squares estimates



Parametric Regression: Example

- Let, $g(x^T|w_1, w_0) = w_1x^T + w_0$ taking derivative of the sum of squared errors

$$\begin{aligned}\sum_t r^T &= Nw_0 + w_1 \sum_t x^T \\ \sum_t r^T x^T &= w_0 \sum_t x^T + w_1 \sum_t (x^T)^2\end{aligned}$$

$$\mathbf{A} = \begin{bmatrix} N & \sum_t x^T \\ \sum_t x^T & \sum_t (x^T)^2 \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} \sum_t r^T \\ \sum_t r^T x^T \end{bmatrix}$$

can be solved $\mathbf{w} = \mathbf{A}^{-1}\mathbf{y}$





CREDITS SOME CS GUY!

