DOI: 10.56452/2021SP-8-018

ISSN: 0974-5823 Vol. 6 (Special Issue 3, November 2021)

International Journal of Mechanical Engineering

Essential Mathematics for Artificial Intelligence (AI) and Machine Learning (ML)

Vrnika Jain[1] and Palki[2]

[1]Student, Department of Computer Science Engineering, Chandigarh Engineering College, Jhanjeri, Mohali-140307, Punjab, India

[2]Chandigarh School of Business, Chandigarh Group of Colleges, Jhanjeri, Mohali-140307, Punjab, India

Abstract: Scientist have been attempting to find out "how we think" and the rationale underlying our thinking, comprehension, and predictions for thousands of years. The field of artificial intelligence advances by attempting to implement that intelligence in machines. It is a current discipline of research and engineering that was founded in 1956 by cognitive scientist Marvin Minsky, where the idea of AI was realised (the future of technology). Machine Learning is a subset of AI that was developed via mathematical modelling of neural networks. This article discusses the mathematics that lies beneath today's AI and ML, as well as the common milestones it has reached thus far.

Keywords: AI, ML, optimization, model.

Introduction

AI is a concept where we can or we try to simulate the human thinking capabilities, his emotions and behavior, whereas, ML is an application or subset of AI which allows machine to learn from data (provided / unprovided) and act accordingly.

Main: AI and ML is besides mathematical operations like addition, subtraction, multiplication and division, where we need to have a good knowledge about algebra, calculus, linear algebra, probability and statistics, information theory, etc. Several systematic conditions and techniques have been used to solve the equations and methods for AI and ML, which can be further used and are used by AI and ML enthusiasts.

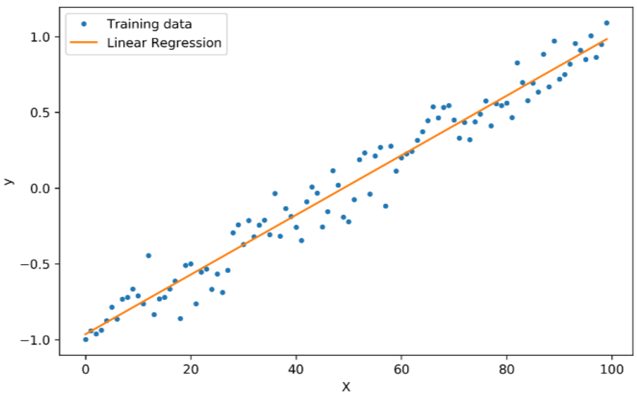
Regression: Regression is a technique for determining the relationship between dependent variables (target) and independent variables (predictor) to predict continuous outcomes.

- Univariate linear regression: A straight line in the form of $h_w(x)= w_1 x+w_0$, where ,
  x is the input,
  h_w is the output,
  w_0 and w_1 are the (weights) real valued coefficients.

For the successful application of linear regression we have to firstly find the values of weights $[w_0, w_1]$ that can furtherly minimize the loss when partially derivative of the equation $\sum_{j=1}^{N} (y_j - (w_1 x_j + w_0))^2$ forms w_0 and w_1 as zero, where y is the variable which will often change with the weights values.

The final solution comes out to be:

$$w_1 = [N(\sum x_j y_j) - (\sum x_j)(\sum y_j)]/[N(\sum x_j^2) - (\sum x_j)^2];$$
$$w_0 = (\sum y_j - w_1(\sum x_j))/N;$$



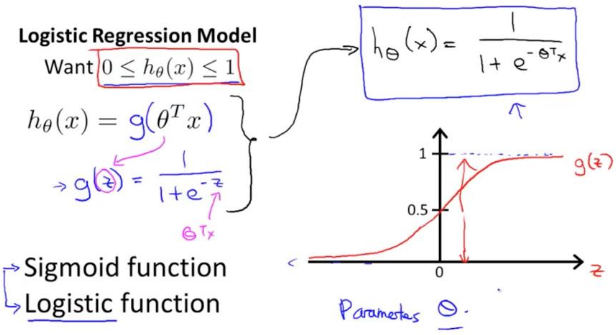
- Multivariate linear regression: An extension of linear regression which is used to predict the dependent variable based on the value of two or more independent variables. It is in the form of $h_{sw}(x_j)=w_0 + w_1x_{j,1} +\ldots\ldots+ w_nx_{j,n} = w_0 + \sum_i w_ix_{j,I}$, but here w_0 term looks left out from the others, and also $x_{j,0} = 1$, therefore, h is dot product of inputs and weights and further forms the equation as:

$$h_{sw}(j) = \sum_I w_ix_{j,I}$$

which has to be minimized using squared loss error loss function reaches one and the updated equation will be:

$$w_i \square w_i+ \alpha \sum_j x_{j,i}(y_j - h_w(x_j))$$

- Logistic Regression: It is used to predict the categorical dependent variables using the given independent values which results probabilistic values i.e. value between 0 and 1.



Here although the two slopes (functions) are very similar in shape, therefore the logistic function is: $h_w(x)=\text{Logistic}(w.x)=1/[1+e^{-w.x}]$, where (w.x) can be said as z for short form.

Artifical Neural Networks: It is simply means building a mathematical model of brain activity (thinking). We all must have read in biology that our brain works properly due to the formation of a electrochemical known as neurons, likewise, in AI and ML, they aims to form those neuron networks in an artificial manner. Each link (connecting two neurons) has a weight $w_{i,j}$ associated with it: $in_j= \sum_{i=0}^{n} w_{i,j}a_i$ which then applies an activation function for the following output :

$$a_j = f(in_j) = f(\sum_{i=0}^{n} w_{i,j}\, a_i)$$

which is sometimes called perceptron or sigmoid perceptron. The error vector $y-h_w(x)$ for any weight's loss is,

$$\delta\, Loss(w)/\, \delta w = \sum_k (\delta/\delta w(y_k - a_k)^2)$$

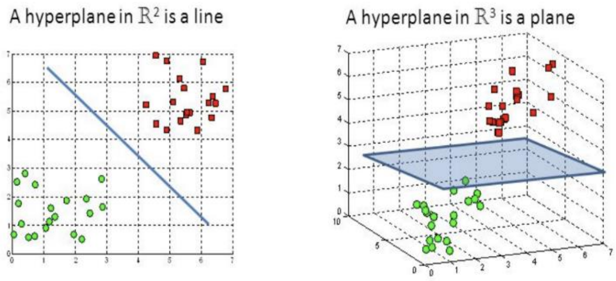
where, the index k ranges over nodes in output layer.

Support Vector Machine: It is also famous by the name SVM which is used to solve classification and Regression problems by determining a hyperplane in an n-dimensional space that can uniquely classify the required data points. The alternative representation which carries out the optimal solution for problem is called alternative representation

$$\sum_j \alpha_j - (1/2) \sum_{j,k} \alpha_j \alpha_k y_j y_k (x_j\, x_k),$$

$$\text{where } \alpha_j >= 0 \text{ and } \sum_j \alpha_j y_j = 0$$

SVMs construct a maximum margin separator which helps the model to generalize well which subsequently forms a decision boundary with largest possible distance till the example points.

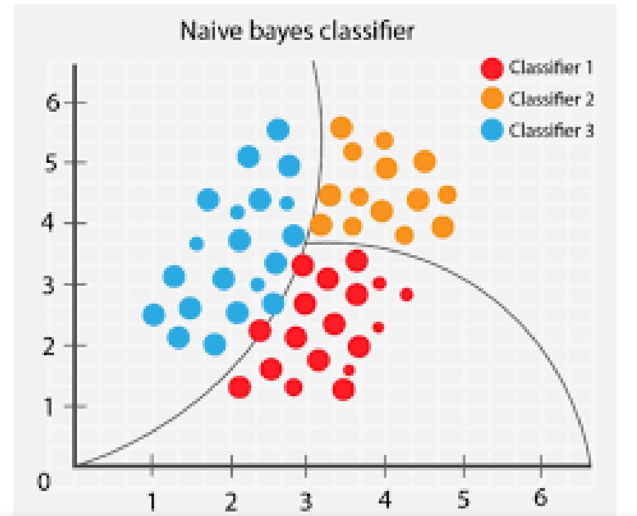


Once the optimal α_j has been calculated, the following property is applicable for the separator also;

$$h(x) = \text{sign}(\sum_j \alpha_j\, y_j\, (x.x_j)\text{-}b)$$

b is to keep intercept as a separate parameter.

<u>Statistical models</u>: It is representation of data and hypotheses which afterwards construct analysis to infer any relationships between independent and dependent variables. Bayesian model is a paradigm for constructing statistical models which is basically based on Baye's Theorem; $p(h_i|d) = \alpha\ p(d|h_i)\ p(h_i)$, where d is the observed value from all the represented data. And the most common Bayesian model used in machine learning is naïve bayes models which is a probabilistic classifier used to solve classification tasks



and the probability of each case is classified as; $p(C|x_1,\ldots,x_n) = \alpha\ p(C)\ \Pi_i\ p(x_i|C)$, here C is the case variable which has to be predicted and $x_1 \ldots x_n$ are the already interposed.

- The continuous models such as Gaussian Naïve Bayes model is an extension of Naïve Bayes in which we substitute the probability density of distribution in the following equation where we need μ as mean and σ as variance for variable x;

DOI: 10.56452/2021SP-8-018

$p(x)=1/(\text{sqrt}.2\,\pi\,\sigma)e^{-(x-\mu)sq/(2\sigma\,sq)}$, the mean and variance

has to be calculated separately here by using the following formulas:

$\mu= \sum_j x_j/N$, and , $\sigma=\text{sqrt}.[(\sum(x_j- \mu)^2/N]$

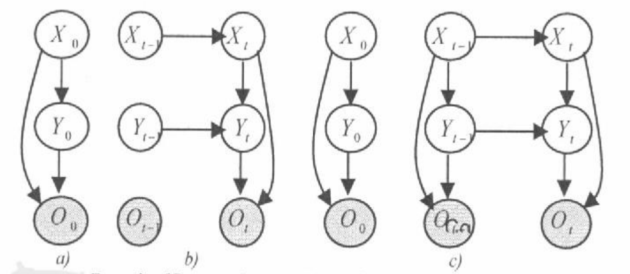
Probabilistic models: These models are the heart of ML and AI. It is a statistical technique used to predict the high probability of occurring a incident as a future outcome by using random events or actions. We can describe probabilistic models into three specific kinds of models:

- Hidden Markov models (HMM): It is used to describe the evolution of factors which are not directly observable in simple words drive the probabilistic characteristic of any random process.

Let X_t be a discrete state variable whose values can be denoted by integers 1,…S, where S is the number of possible states. The transition model $p(X_t|X_{t-1})$ becomes SxS matrix T, where $T_{ij} = p(X_t=j \mid X_{t-1}=i)$, T_{ij} is the probability of transition from i to j state. Here to solve a transition we usually put sensor model in matrix form where e_t is the evidence variable at time t, needed to be specified for each state using $p(e_t|X_t=i)$ for each state i keeping the (O_t), i^{th} diagonal entry $p(e_t|X_t=i)$ and other values 0.
After using column vectors, the forward and backward equations come out to be;

$$f_{1:t+1}= \alpha\, O_{t+1}\, T^T f_{1:t} \text{ and } b_{k+1:t} = TO_{k+1}b_{k+2:t}$$

- Kalman Filters: The optimal estimator with linear dynamic systema and Gaussian noise is the Kalman filter: $p(X_{t+\Delta} = x_{t+\Delta} \mid X_t=x_t\ ,\ X_t = x_t)=N(x_t+x_t\ \Delta,\sigma^2)(x_{t+\Delta})$, where Δ is the time interval between observations, constant velocity during interval, and $X_{t+\Delta} = X_t + X\Delta$. And this equation can be further updated according to the problem.
- Dynamic Bayesian Networks (DBN): DBN is a bayesian network which is used to relate different variables to each other in an adjacent time laps. To construct a DBN, there are three basic information required; $p(X_0)$: distribution over state variable, $p(X_{t+1}|X_t)$: transition model, $p(e_t|X_t)$: sensor model.



Example of Dynamic Bayesian Network a) Prior Network b) Transition Network c) Dynamic Bayesian Network: Combination of prior and transition networks

Application: AI and ML are becoming a significant staple of innovation. Computer scientists and mathematicians are using AI and ML to help them to solve and suggest new mathematical theorems in the most complex field i.e. knot theory and representation theory. ML is basically used here to assist all the work of analysis of complex data sets and suggest possible lines of attack for unproven ideas in mathematics.

REFERENCES:

[1]. Artificial Intelligence: A Modern Approach, Stuart J. Russell, Peter Norvig, 2016
[2]. Deep Learning and Mathematical Intuition: A Review of (Davies et al. 2021)
[3]. Quantum Mathematics in Artificial Intelligence
[4]. A Timeline in Machine Learning History, Wesley Chai
[5]. All the Math you need to know in Artificial Intelligence, Jason Dsouza
[6]. Google AI system proves over 1200 mathematical theorems, David H Bailey
[7]. Bayesian machine learning, DataRobot, 2020
[8]. Maths researchers hail breakthrough in applications of artificial intelligence, University of Sydney
[9]. Gaussian Naïve Bayes, Rohan Vats