DOI: 10.56452/2021SP-8-019

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

International Journal of Mechanical Engineering

Mathematics behind Artificial Intelligence and Machine Learning

Rashmi Choudhary[1,*] and Dinanter Pal Kaur[2]

Department of Applied Science, Chandigarh Engineering College, Jhanjeri, Mohali-140307, Punjab, India

[2] Department of Agriculture, Chandigarh Engineering College, Jhanjeri, Mohali-140307, Punjab, India

E-mail: rashmi.j1543@cgc.ac.in

Abstract: Since thousands of years ago, we have been attempting to comprehend "how we think" and the reasoning behind our perceptions, understanding, and predictions. By attempting to incorporate that intellect into technology, artificial intelligence advances. The idea of artificial intelligence (AI, the future of technology) was realised in this contemporary field of research and engineering, which was founded in 1956 by cognitive scientist Marvin Minsky. The branch of artificial intelligence known as machine learning (ML) was initially developed from the mathematical modelling of neural networks. This paper explains the mathematics that underlies modern AI and ML as well as the shared milestones that have been reached thus far.

Keywords: Artificial Intelligence, Machine Learning, Optimization model.

Introduction

Artificial intelligence (AI) is a concept where we can or attempt to replicate human thinking abilities, emotions, and behaviour, whereas machine learning (ML) is an application or subset of AI that enables machine to learn from data (given or unprovided) and act accordingly. Through computer programming, conventional techniques, and clear instructions, there are many activities and problems that are highly challenging to tackle in the world. Making computer games, iPhone apps, or desktop applications is relatively achievable using conventional methods, but creating a machine that can outperform the greatest human in a challenging game or a vehicle that can drive itself are far more difficult.

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 ,

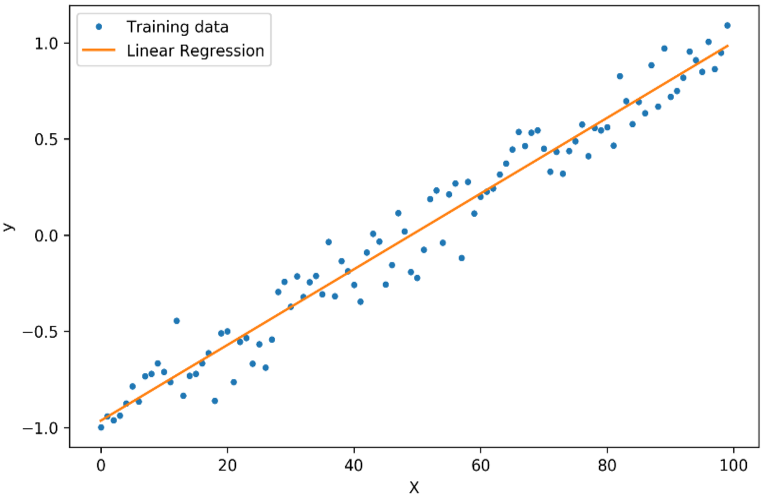
  x is the input,

  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: $w1 \ = \ [N(\sum xjyj) - (\sum xj)(\ \sum yj)]/[N(\sum \ xj2) - (\sum xj)2];$

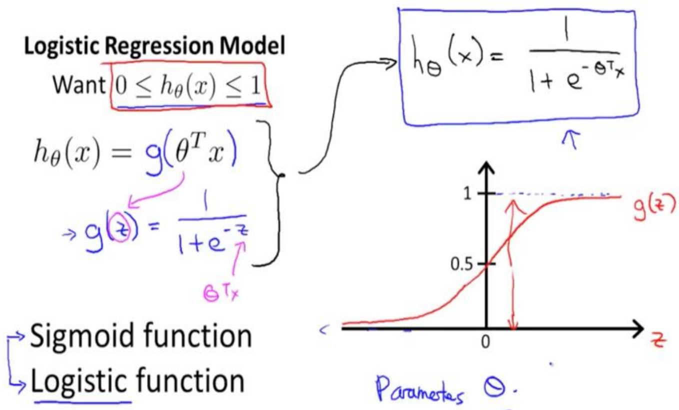
$w0 \ = \ (\sum yj \ - \ w1(\sum xj))/N;$



- **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 $hsw(xj) = w0 \ + \ w1xj,1 \ + \cdots \ldots . + wnxj,n \ = \ w0 \ + \ \sum iwixj$,I, but here w_0 term looks left out from the others, and also $xj,0 \ = \ 1$, therefore, h is dot product of inputs and weights and further forms the equation as: $hsw(j) \ = \ \sum Iwixj,I$,which has to be minimized using squared loss error loss function reaches one and the updated equation will be:

$$wiwi + \alpha \sum jxj,i(yj - hw(xj))$$

- **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: $hw(x) = Logistic(w.x) = 1/[1 + e - w.x]$, $where \ (w.x)$ can be said as z for short form.

2.**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 $wi, j\ associated\ with\ it: inj = \sum i = 0\ nwi, jai\ which\ then\ applies\ an\ activation\ function\ for\ the\ following\ output:$

$$aj = f(inj) = f(\sum i = 0\ nwi, j\ ai)$$

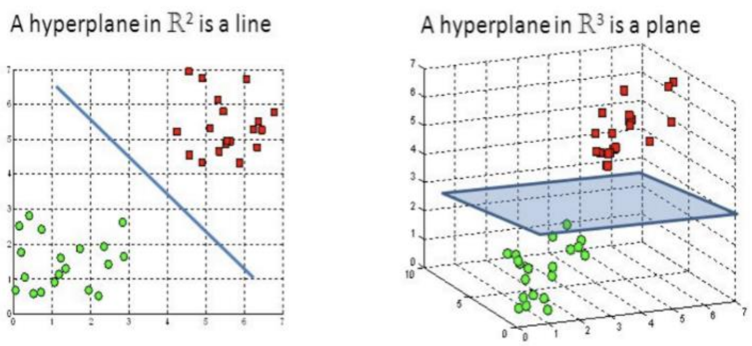
which is sometimes called perceptron or **sigmoid perceptron**. The error vector $y - hw(x)$ for any weight's loss is,

$\delta\ Loss(w)/\ \delta w = \sum k\ (\delta/\delta w(yk - ak)2)$, 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 problemis called alternative representation

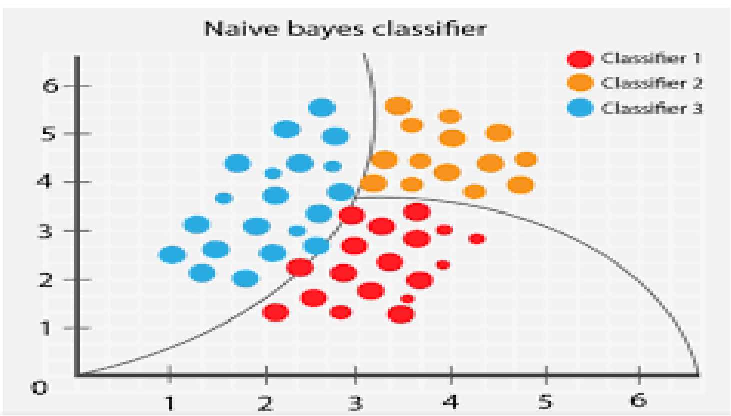
$$\sum j\alpha j - (1/2) \sum j, k\alpha j\alpha kyjyk\ (xjxk), where\ \alpha j >= 0\ and\ \sum j\alpha jyj = 0$$

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;
$h(x) = sign(\sum j\alpha jyj\ (x.xj) - b)$, $b\ is\ to$ keep intercept as a separate parameter.

Statistical models: 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 $Baye's\ Theorem;\ p(hi|d) = \alpha\ p(d|hi)\ p(hi), where$ 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



DOI: 10.56452/2021SP-8-019

and the probability of each case is classified as; $p(C|x1, \dots, xn) = \alpha\, p(C)\, \Pi i\, p(xi|C), here\ C\ is$ the case variable which has to be predicted and $x1 \dots xn$ 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;

$$p(x) = 1/(sqrt. 2\, \pi\sigma)e - (x - \mu)sq/(2\sigma sq),$$

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

$$\mu = \sum jxj/N\ , and\ , \sigma = sqrt. [(\sum(xj - \mu)2/N]$$

<u>Probabilistic models</u>: 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:

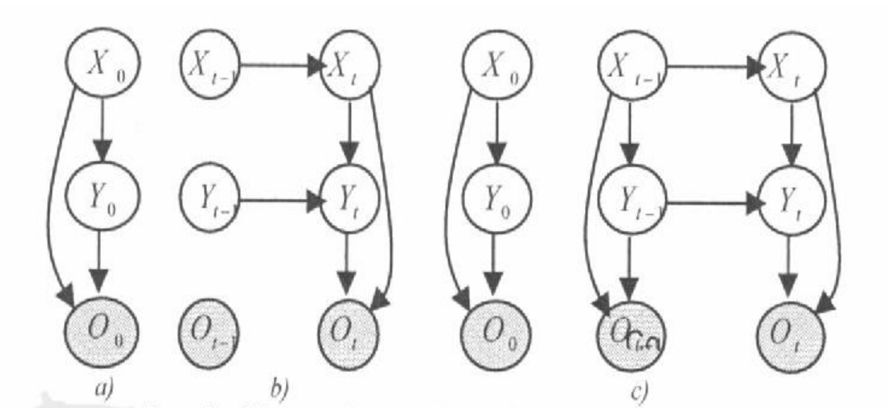
- **<u>Hidden Markov models (HMM)</u>**: It is used to describe the evolution of factors which are not directly observable in simple words drive the probabilistic characteristic of any random process.

Let $Xt\ be$ a discrete state variable whose values can be denoted $by\ integers\ 1, \dots S, where\ S$ is the number of possible states. The transition model $p(X_t|X_{t-1})$ becomes SxS matrix T, $where\ Tij = p(Xt = j \mid Xt - 1 = i), Tij\ is$ 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 $p(et|Xt = i)\ for$ each state $i\ keeping\ the\ (Ot), ith$ diagonal entry $p(et|Xt = i)\ and\ other$ values 0.
After using column vectors, the forward and backward equations come out to be;

$$f\ 1{:}\,t + 1 = \alpha\ Ot + 1\ TTf1{:}\,t\ and\ bk + 1{:}\,t\ =\ TOk + 1bk + 2{:}\,t$$

- **<u>Kalman Filters</u>: The** optimal estimator with linear dynamic systema and Gaussian noise is the Kalman filter: $p(Xt + \Delta\ =\ xt + \Delta \mid Xt = xt\ , Xt\ =\ xt) = N(xt + xt\Delta, \sigma 2)(xt + \Delta)$, where Δ is the time interval between observations, constant velocity during interval, $andXt + \Delta\ =\ Xt\ + X\Delta$. And this equation can be further updated according to the problem.

- **<u>Dynamic Bayesian Networks (DBN)</u>: 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, $p(Xt + 1|Xt)$: transition $model, p(et|Xt){:}\ sensor\ model.$



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.

DOI: 10.56452/2021SP-8-019

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
10] . Le, DN., Parvathy, V.S., Gupta, D. et al. IoT enabled depthwise separable convolution neural network with deep support vector machine for COVID-19 diagnosis and classification. Int. J. Mach. Learn. & Cyber. (2021)
11]. Sankhwar, S., Gupta, D., Ramya, K.C. et al. Improved grey wolf optimization-based feature subset selection with fuzzy neural classifier for financial crisis prediction
12] Zhenru Wang, Tijie Shi, " Prediction of the Admission Lines of College Entrance Examination based on machine learning", 2nd IEEE International Conference on Computer and Communications",
13] Aaron N. Richter,b,∗, Taghi M. Khoshgoftaar, "A review of statistical and machine learning methods for modeling cancer risk using structured clinical data", Artificial Intelligence In Medicine
14] Mohan S Acharya, Asfia Armaan, Aneeta S Antony, "A Comparison of Regression Models for Prediction of Graduate Admissions", Second International Conference on Computational Intelligence in Data Science, 2019, IEEE
15] Somya Goel, Sanjana Rosshan, Rishabh Tyagi, Sakshi Agarwal, "Augur Justice: A Supervised Machine Learning to predict Outcomes of Divorce court cases", Fifth International Conference on Image Information Processing"