Applications of Exponential Functions in Artificial Intelligence

-Risiraj Dey, Dr. Anitha N

Exponential functions have always been useful in physics, economics, chemistry, engineering and various other domains.

The first use of the exponential function was in calculating the compound interest , where "a quantity grows or decays at a particular rate proportional to its current value". This was discovered by Jacob Bernoulli in 1683.In fact he found e^(1) or e.

The total debt owed by a person to a bank can be given by:

$$P(1+r)^n$$

Where P is the initial amount of money borrowed, r is the rateof interest and n is the no. of times the interest is compounded.

To make things a bit simpler, lets assume P as 1, and while we reduce r to 1/n where n approaches infinity, the amount converges to e which is 2.718281828...:

$$1 \times (1 + 1/n)^n$$

Leonhard Euler first stated the function : e^(x), which is {e*e*e.....}

The exponential can be defined as a limit of a function using binomial theorem in the following manner.

$$\exp x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n.$$

Here n is tending towards infinity. It is clear that as x takes higher values, the value of e changes rapidly to much higher values.

Its interesting to note that the derivative and integral of the exponential function are the same.

$$\frac{d}{dx}e^x = e^x \log_e e = e^x.$$

ISSN 2229-5518

It follows from the fact how real numbers raised to another real number can be expressed mathematically. The logarithmic function, being its inverse (ie e^(-1)) makes the derivative of exponential being equal to itself. Integrating both sides would also make no difference.

This property of exponential makes it particularly useful in solving differential equations.

In fact, many differential equations give rise to exponential functions, including the Schrödinger equation and the Laplace's equation as well as the equations for simple harmonic motion.

No doubt the exponential functions have helped in the growth of Mathematics and various branches of Science, but it has a long history attached to it.

Whenever a mathematician is struck in finding a solution to a problem, which of course is not solvable in trivial ways, it is the exponential function that has come to the rescue.

Geometry of Exponential Function:

The natural logarithm can be defined as:

$$\ln a = \int_1^a \frac{1}{x}\, dx.$$

we see that the area between the "standard" hyperbola **xy=1** and the horizontal axis between **1** and **x** is log(x).

So from this we get **a** = **exp(x)** which is the vertical line such that area between **a** =**1** and **a** = **exp(x)**, between the hyperbola and x axis is **x**.

While dealing with differential equations one has to various transformations in order to find solutions. One of them is the **Laplace transform.** Formally stated as:

$$F(s) = \int_0^\infty f(t)e^{-st}\, dt$$

which involves the exponential function multiplied with the function to be transformed.

Now that we are clear about why the exponential is used to solve differential equations lets now see what makes the Laplace Transform to be a goto approach for solving them:

The differential equations are actually transformed into algebraic equations, which are much easier to deal with. So if the function f(t) say e^(a*t) has poles at say s1,s2,s3.. and so on, then f(t) has time domain behaviour of the form e^(s1*t), e^(s2*t),...e^(sn*t). The decaying exponential decays faster than f grows so the analysis of the function in the transformed domain becomes easier.

The Fourier transform is just a generalized case of the Laplace transform, the difference being that a complex exponential is used rather the real valued one and the limits vary to -infinity to +infinity.

In every manner the exponential function proves to be very helpful in almost every field where mathematical rigour is involved.

It is seen that the exponential acts as a kernel in Laplace and Fourier Transforms. Special functions like the Beta, Gamma, Bessel functions are also defined using the exponential.

Artificial Intelligence and Statistics:

In Neural networks and especially in Deep learning, the logistic function (sigmoid activation function) is used because they introduce non linearity; meaning that if combinations of linear functions are used, there would be completely different non-linear models making it difficult to generate accurate results.

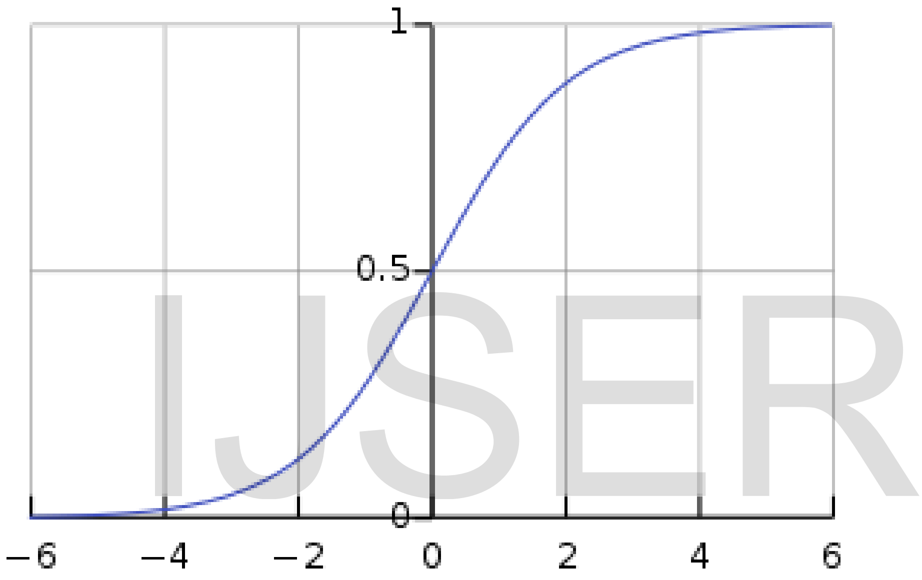
It's smooth and differentiable nature makes it easier to derive the back propagation algorithm. It is also monotonic in its behaviour.

The sigmoid function is found to compress the outputs of neurons in multi layer neural nets.

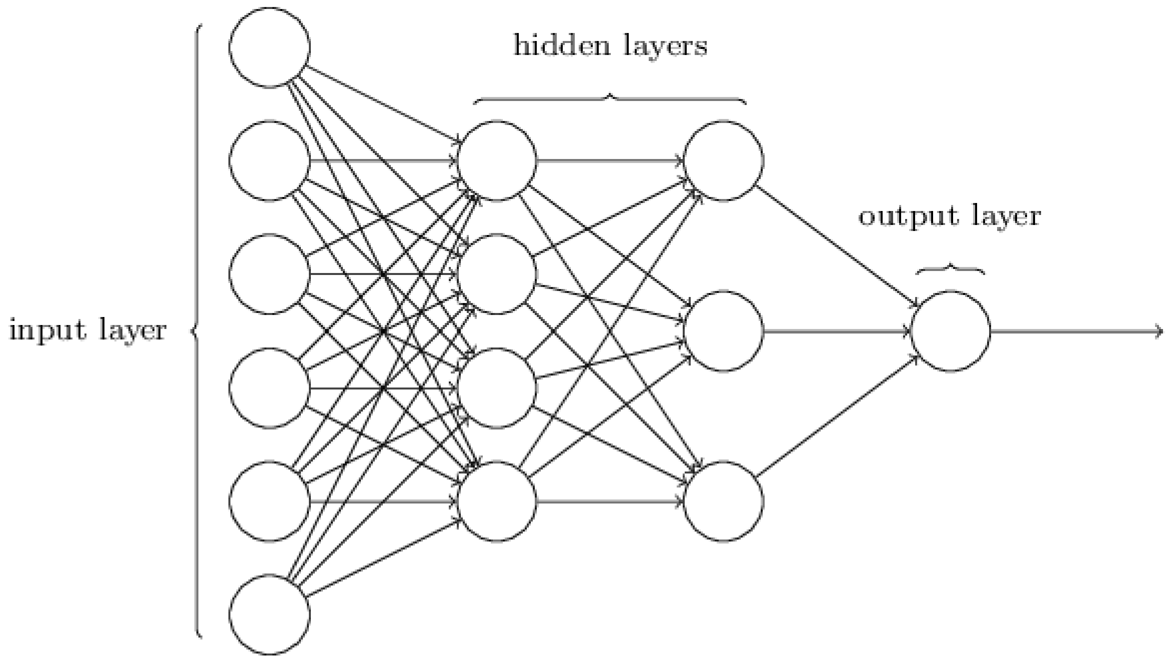
The sigmoid function:

$$f(x) = \frac{1}{1 + e^{-x}}$$

x being the sum of weight multiplied with input data which is summed with a bias value.



A Multi-Layer NN:



The sigmoid function has multiple optimal solutions and can be easily used for problems which progress from a small initial beginning to undergo a change(accelerate) to reach some point. Multiple optima exist as the function is convex for values < 0 and concave for values >0.

Several kinds of sigmoid functions including the logistic and hyperbolic tangent functions have been used as the activation function of artificial neurons. Sigmoid curves are also common in statistics as cumulative distribution functions (which go from 0 to 1), such as the integrals of the **logistic distribution, the normal distribution, and Student's *t* probability density functions.**

Some of the examples of functions which are sigmoidal in nature are:

The hyperbolic tangent:

$$f(x) = \tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

The error function: (particularly useful in cumulative distribution function of a normal distribution)

$$f(x) = \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt$$

Another useful function is the softmax function which is also known as the generalized exponential function. It's mainly used for distributing the outputs of Neural nets to a probability distribution over predicted output classes. So basically probabilistic models can be generated. Such models are used for **NLP(Natural Language Processing).**

The softmax function is defined as:

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \text{ for } i = 1, \dots, K \text{ and } \mathbf{z} = (z_1, \dots, z_K) \in \mathbb{R}^K$$

where z is the input vector into which the exponential is applied and then normalized by dividing them by the sum of exponentials upto the kth exponential.

In the area of Reinforcement learning the softmax function can convert values into probability outcomes.

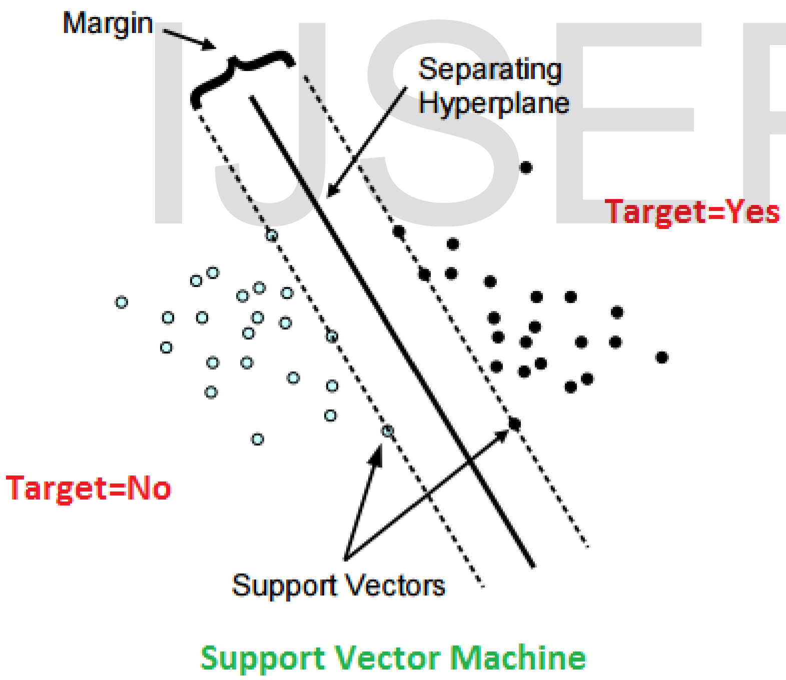
In Logistic regression/maximum entropy Classifiers,
exponential functions make the optimization objective convex
which will ensure global minimum for the classification error function
which is to be minimized.

Of the various non-linear algorithms the, **Gaussian Naive Bayes Classifier** can extend to many numerical attributes by assuming a Gaussian distribution.
The function used for prediction is given by the Gaussian Probability Distribution Function:

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

In purely mathematical terms, this is the normal distribution.

SVM(Support Vector Machines) are implemented using various kernels, and with each variation, the measure between new data and the support vectors is defined.



A popular kernel function used in support vector machine classification is the radial basis function kernel or RBF kernel. Considering two samples x, y the RBF kernel is represented as:

$$k(x,y)=\exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$$

Both of the above stated functions use the exponential function on similar terms.