Linear Regression

Features

The features are the elements of your input vectors. The number of features is equal to the number of nodes in the input layer of the network

Category	Features
Housing Prices	No. of Rooms, House Area, Air Pollution, Distance from facilities, Economic Index city, Security Ranking etc.
Spam Detection	presence or absence of certain email headers, the email structure, the language, the frequency of specific terms, the grammatical correctness of the text etc.
Speech Recognition	noise ratios, length of sounds, relative power of sounds, filter matches
Cancer Detection	Clump thickness, Uniformity of cell size, Uniformity of cell shape, Marginal adhesion, Single epithelial cell size, Number of bare nuclei, Bland chromatin, Number of normal nuclei, Mitosis etc.
Cyber Attacks	IP address, Timings, Location, Type of communication, traffic details etc.
Video Recommendations	Text matches, Ranking of the video, Interest overlap, history of seen videos, browsing patterns etc.
Image Classification	Pixel values, Curves, Edges etc.

Weights

Weights correspond to each feature.

Weights denote how much the feature matters in the model.

Higher weight of a particular feature means that it is more important in deciding the outcome of the model.

Weights of a feature represent that how much evidence it gives in favor or against the current hypothesis in context of the existence or non-existence of the pattern you are trying to identify in the current input.

Generally weights are initialized randomly.

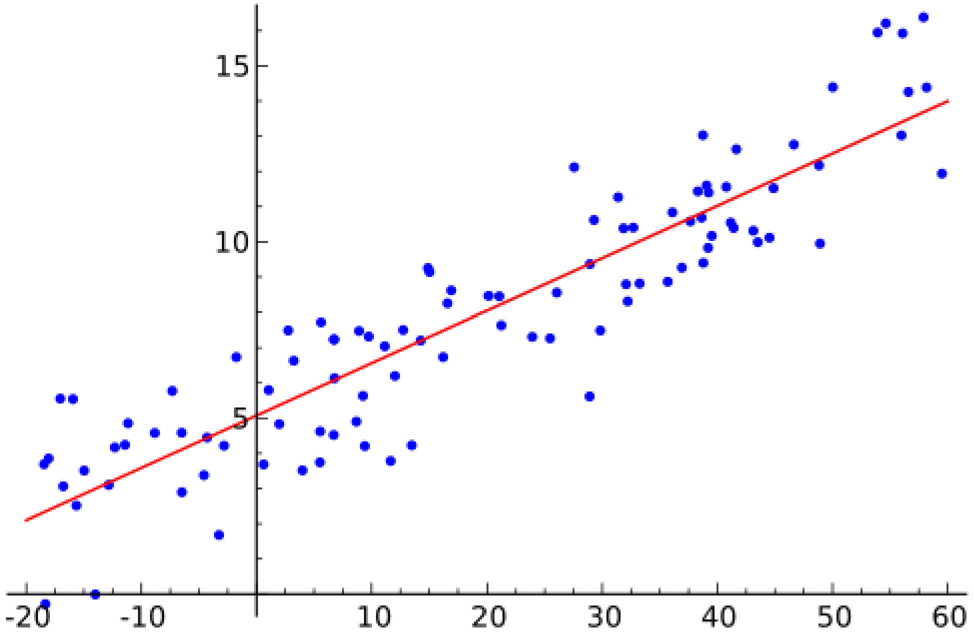
we try to bring them to near optimal values so that they are able to fit the model well and can help in prediction of unseen values

Linear Regression and Logistic Regression

Linear Regression: For applications where output will be a real value e.g. Predicting housing price, or predicting price of a share in stock market. In most cases we have multiple dependent variables, and we call it multiple linear regression

Logistic Regression: For applications where the output will be a binary value (0/1). E.g. whether this Medical Image depicts Tumor or not

Linear Regression



Dependent Variable

Indendent Variable

Where line crosses the y-axis

$$y = m\,x + b$$

Coeficient, Rate and Slope of line

Y- Intercept

$$\mathrm{MSE} = \frac{1}{m}\sum_{i=1}^{m}(\hat{Y}_i - Y_i)^2$$

Benefits of squaring

Squaring always gives a positive value, so the sum will not be zero.

Squaring emphasizes larger differences—a feature that turns out to be both good and bad (think of the effect outliers have).

Linearity Vs. Non-Linearity

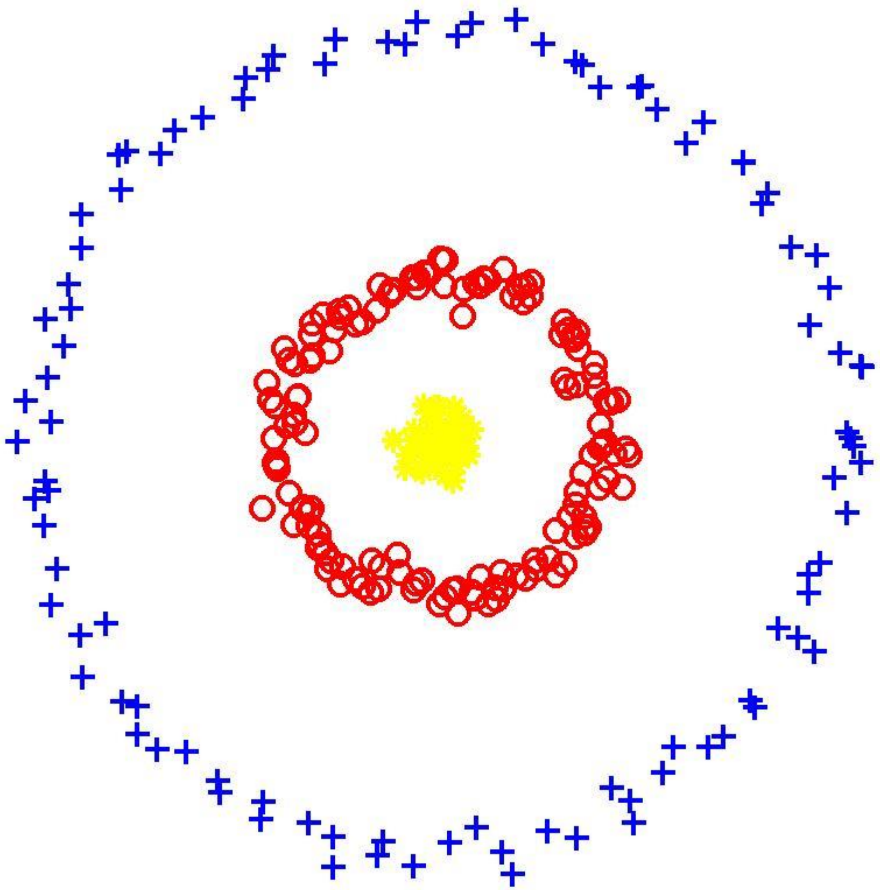
Simple multiplication of weights with inputs and giving the outputs will be linear function.

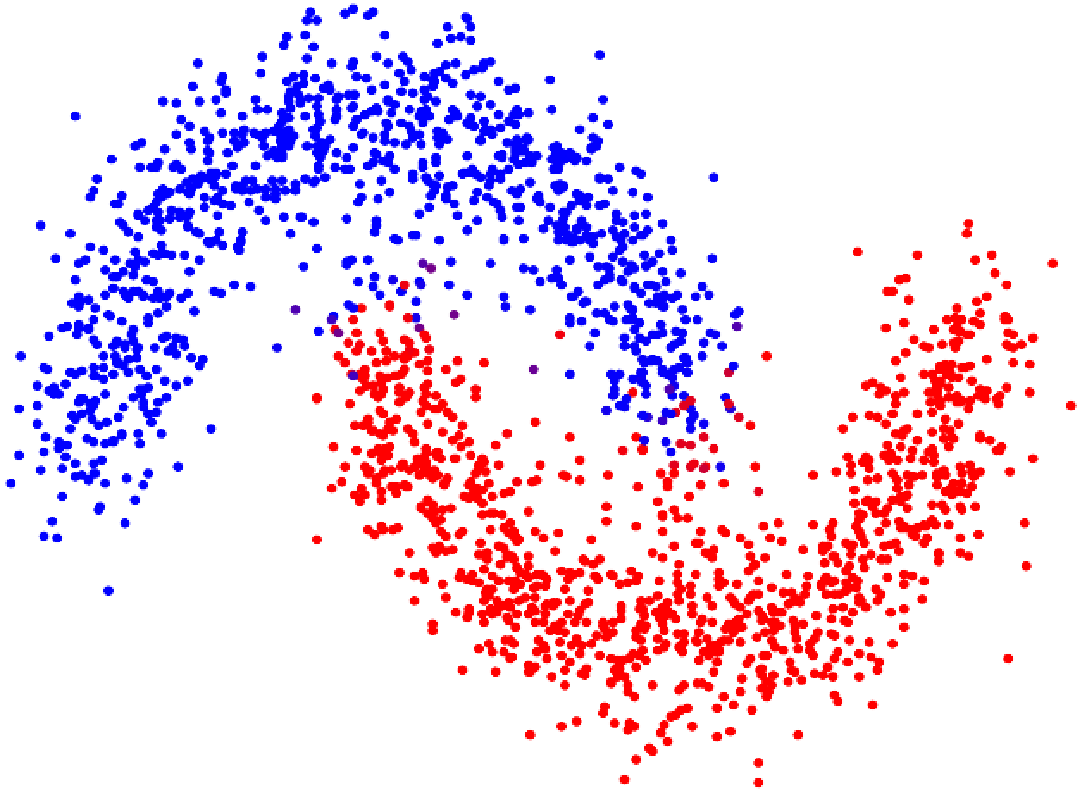
A linear function is just a polynomial of one degree and will not be able to learn complex functions.

Linear functions don't have much expressive power and will loose out when solving complex problems.

Without activation functions (which will help us to achieve non-linearity) Neural Network will not be able to learn unstructured data like images, audio data, videos and text.

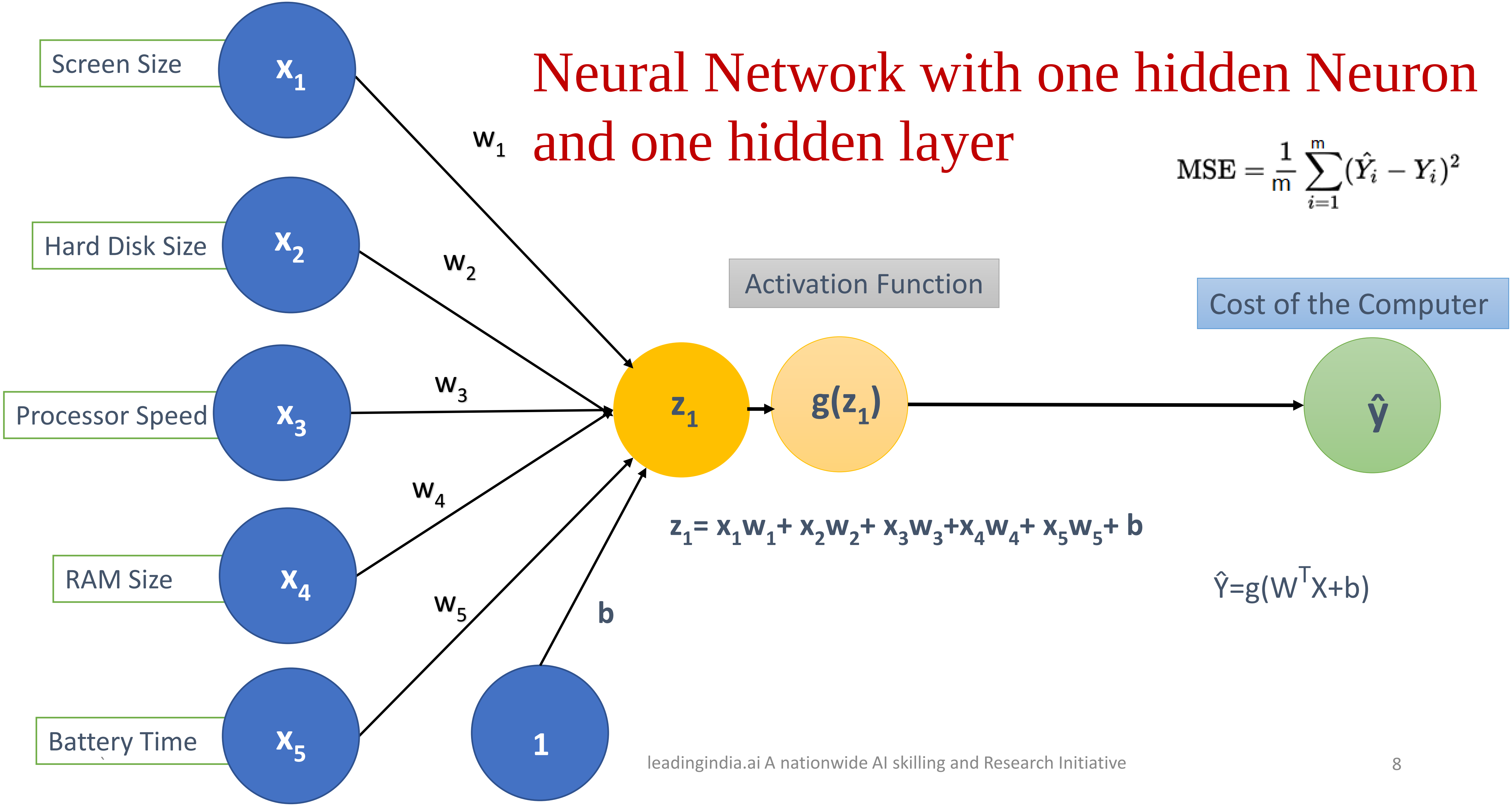
Non Linear Patterns





Neural Network with one hidden Neuron and one hidden layer

$$\text{MSE} = \frac{1}{m}\sum_{i=1}^{m}(\hat{Y}_i - Y_i)^2$$



$$z_1 = x_1w_1 + x_2w_2 + x_3w_3 + x_4w_4 + x_5w_5 + b$$

$$\hat{Y} = g(W^T X + b)$$

Why W^TX

$$\begin{bmatrix} x_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{bmatrix} \quad \begin{bmatrix} W1 \\ W2 \\ w_3 \\ w_4 \\ w_5 \end{bmatrix}$$

Matrix Size: (5X1)

Matrix Size: (5X1)

Multiplication of these two matrices is not possible

$$\begin{bmatrix} W1 & W2 & W3 & W4 & W5 \end{bmatrix} \quad \begin{bmatrix} x_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{bmatrix}$$

Matrix Size: (1X5)

Matrix Size: (5X1)