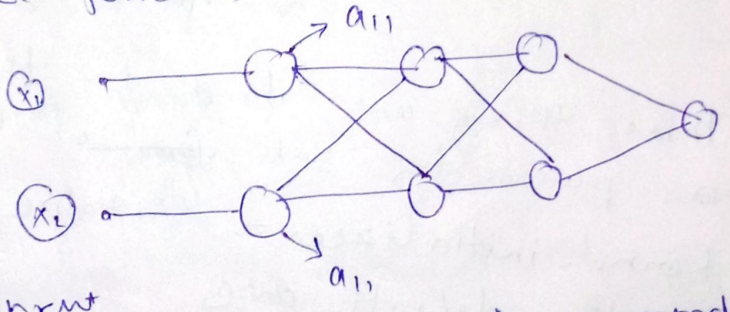
05 Oct - 2024

BATCH NORMALIZATION

- it is a technique that is used to speed up the train speed of the our neural network.

What is Batch Norm?

- Batch Normalization (BN) is an algorithm method which makes the training of Deep Neural Network. (DNN) faster and more stable.
- it consists of normalizing activation vectors from hidden layers using the mean and variance of the current batch. This normalisation step is applied right before (or right after) the non-linear function.

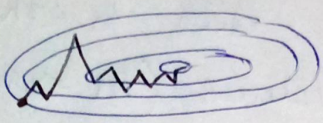
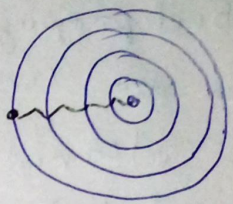


input

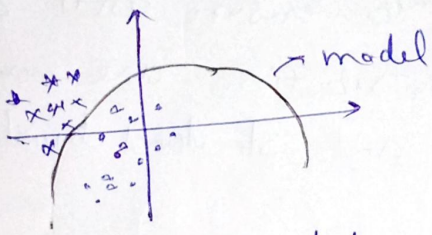
here x_1, x_2 is normalized input.
mean = 0 . sd = 1

and also in batch normalization normalize the activation function also a_{11}, we do this for all hidden layer in the neural network.

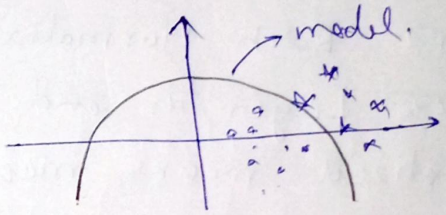
why use Batch Normalization ?

if input data is not normalized.	if input data is normalized.
 contour plot training slow	 contour plot training is faster as well as faster

~~internal~~ Covariate shift:



on training data

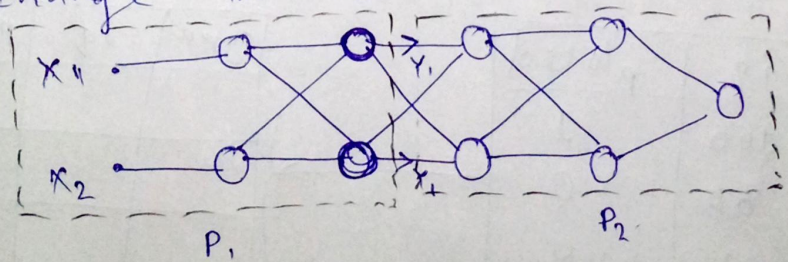


on testing data.

even a the relation between the x and y are same. in both case, retrain the model.

internal covariate shift:

we define internal covariate shift as the change in the distribution of network activations due to change in network parameters during training.



- here P_1 and P_2 are two neural network connected to each other.

- Y_1 and Y_2 are output of neural network P_1 but Y_1 and Y_2 are input for neural network P_2
- but in the ~~network~~ whole neural network. (Backpropogation) in internal weights and biases are continuously change. due to this the internal distributions values (for example. Y_1 and Y_2) continuously change and that's result the our neural network become unstable
- to handle this situation in neural network we apply Batch normalization.
- ~~इसमें~~ Batch Normalization यह ensure करता है for ~~हर~~ Layer के and ~~में~~ distribution are ~~not~~ normalized means. mean = 0 and standard deviation = 1

things that we keep in mind while using. Batch Normalisation:

- it is apply with. Mini-Batch Gradient descent
- it applies layer by layer

How Batch Normalization Works:

cgpa	iq	placed
8.9	100	1
6.2	89	0
9.1	91	0
7.7	76	1

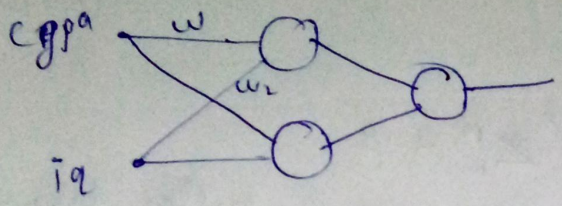
input raw data

cgpa	iq	placed

Batch size = 4

input normalized data

let batch size = 4



$$Z_{11} = w_1 \text{cgpa} + w_2 \text{iq} + b$$

Here there is two ways to normalized the distributions for each layer.

(i) $Z_{11} \longrightarrow Z_{11}^N \longrightarrow g(Z_{11}^N) = a_{11}$

(ii) $Z_{11} \longrightarrow g(Z_{11}) \longrightarrow a_{11} = a_{11}^N$

the first one is mostly used. approach

$$\frac{Z_{11} - \mu}{\sigma} = Z_{11}^N$$

$\mu \rightarrow$ mean $\quad \sigma$ – standard deviation.

mean is calculated for each batch differently

$$\mu_B = \frac{1}{m} \sum_{i=1}^{m} Z_{11}^i \qquad \sigma_B = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (Z_{11}^i - \mu_B)}$$

here $m = 4$

calculate μ_B and σ_B for each ~~batch~~ neuron differently.

then for every neuron. $\varepsilon \rightarrow$ error term

$$Z_{11}^i = \frac{Z_{11}^i - \mu_B}{\sigma_B + \varepsilon}$$

then all value come under the range of 0-1

and $\mu = 0$, $\sigma = 1$

$$z_{11} \rightarrow z_{11}^{N}$$

$$z_{11}^{BN} \rightarrow \gamma z_{11}^{N} + \beta$$

here γ and β learnable parameter like w, b
the initial value of γ and β is 1 and 0 respectively.

$$z_{11} \rightarrow z_{11}^{N} \rightarrow z_{11}^{BN} \rightarrow g(z_{11}^{BN}) = a_{11}$$

the above method is apply for each neuron in neural network. and each neuron has it's own γ and β value

Batch Normalization step:

→ Batch Normalize
→ Scale and shift (γ and β)

- in neural network. Batch Norm can be considere as a layer. after each layer.
- γ and β are update also using gradient descent
- each neuron store four value 2 is learnable (γ and β) and 2 is unlearnable. (μ, σ)

Advantages:

- make training more stable.
- Training becomes faster.
- Batch normalization can act as regularizer
- reduce the impact of weight initialization.