NPTEL Week 7 Live Session

Powered by:

on Deep Learning (noc24_ee04)

A course offered by: Prof. Prabir Kumar Biswas, IIT Kharagpur

- Week 6 quiz solution (Autoencoder, PCA)
- Week 7 practice questions (Variety of AE, Convolution, Cross-correlation)

PMRF
Prime Minister's Research Fellows
Ministry of Education
Government of India

By

Arka Roy
NPTEL PMRF TA

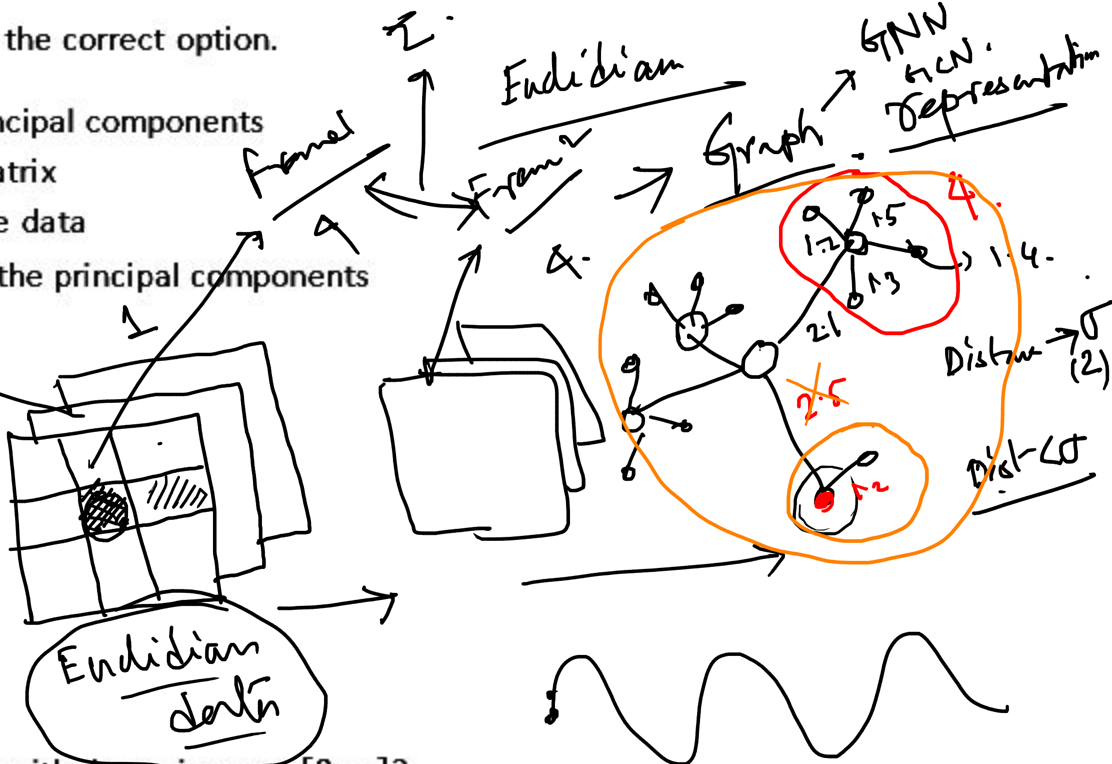
Prime Minister's Research Fellow
Department of Electrical Engineering, IIT Patna
Web: https://sites.google.com/view/arka-roy/home

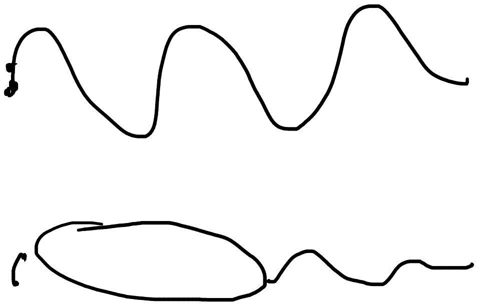
Which of the following is not true for PCA? Choose the correct option.

a. Rotates the axes to lie along the principal components
b. Is calculated from the covariance matrix
c. Removes some information from the data
d. Eigenvectors describe the length of the principal components

(i) R peak dist.
(ii) T wave timing.
(iii) shape of pulse
(iv) ST segment.

RGB.





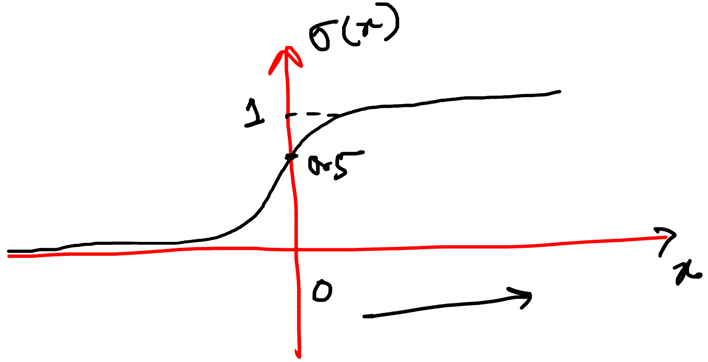
What is the output of sigmoid function for an input with dynamic range $[0,\infty]$?

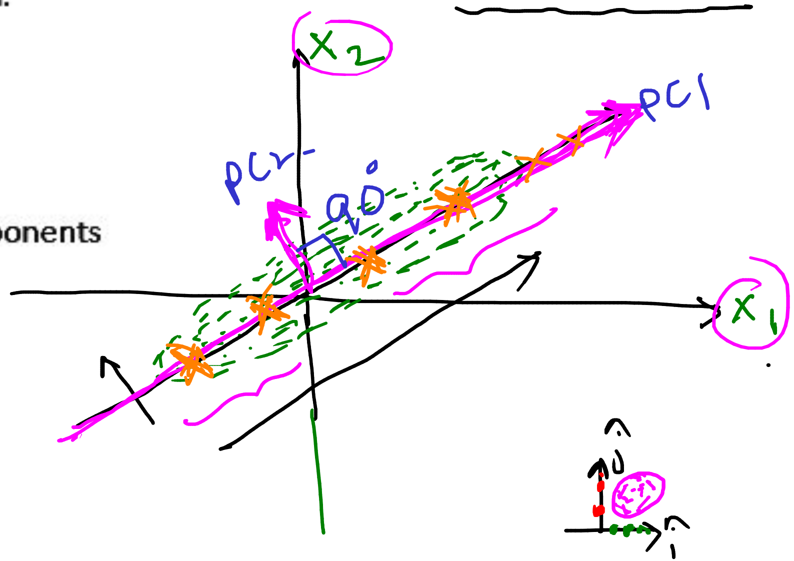
a. $[0,1]$
b. $[-1,1]$
c. $[0.5,1]$
d. $[0.25,1]$

WEEK-6 Quiz

1 Which of the following is not true for PCA? Choose the correct option.

a. Rotates the axes to lie along the principal components
b. Is calculated from the covariance matrix
c. Removes some information from the data
d. Eigenvectors describe the length of the principal components





What is the output of sigmoid function for an input with dynamic range $[0, \infty]$?

a. $[0,1]$
b. $[-1,1]$
c. $[0.5,1]$
d. $[0.25,1]$

A zero-bias autoencoder has 3 input neurons, 1 hidden neuron and 3 output neurons. If the network is perfectly trained using an input $\begin{bmatrix}2\\3\\5\end{bmatrix}$. What would be the values of the weights in the autoencoder?

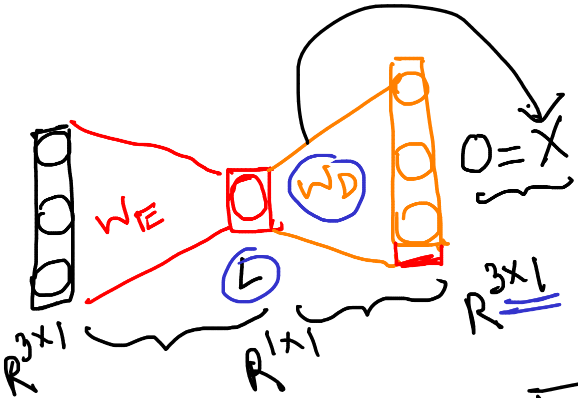
a. $[1 \quad 1 \quad 1], \begin{bmatrix}2\\3\\5\end{bmatrix}$

b. $[1 \quad 1 \quad 1], \begin{bmatrix}0.2\\0.3\\0.5\end{bmatrix}$

c. $[0.2 \quad 0.3 \quad 0.5], \begin{bmatrix}1\\1\\1\end{bmatrix}$

d. $[2 \quad 3 \quad 5], \begin{bmatrix}1\\1\\1\end{bmatrix}$

$X = \begin{bmatrix}2\\3\\5\end{bmatrix} \Rightarrow$



$\mathcal{L} = \|O - X\|^2$

$= MSE(O, X)$. very low.

$= 0$

$L \, (\in R^{1\times1}) = [W_E]_{1\times3} \begin{bmatrix}2\\3\\5\end{bmatrix}_{3\times1}$

$L = W_E^t X$

option (a), $L = [1 \; 1 \; 1]\begin{bmatrix}2\\3\\5\end{bmatrix} = 2+3+5 = 10$

$O = \begin{bmatrix}2\\3\\5\end{bmatrix} \cdot 10 = \begin{bmatrix}20\\30\\50\end{bmatrix}$

$\underset{(3\times1)}{O} = \underset{3\times1}{W_D} \cdot \underset{1\times1}{L}$

(b) $L = [1 \; 1 \; 1]\begin{bmatrix}2\\3\\5\end{bmatrix}$

$= 10.$

$O = \begin{bmatrix}0.2\\0.3\\0.5\end{bmatrix} \times 10 = \begin{bmatrix}2\\3\\5\end{bmatrix}$

$O = X$

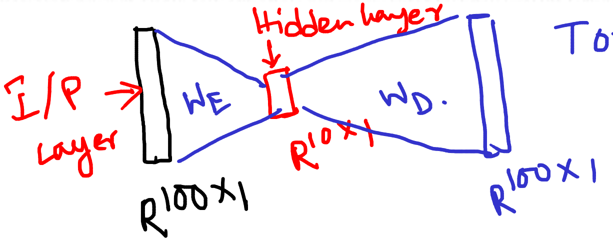
A single hidden and no-bias autoencoder has 100 input neurons and 10 hidden neurons. What will be the number of parameters associated with this autoencoder?

a. 1000

b. 2000

c. 2110

d. 1010



Total Parameter

$W_E: 100 \times 10 = 1000$

$W_D: 10 \times 100 = 1000$

2000

$y = \min(a, b)$ and $a > b$. What is the value of $\frac{dy}{da}$ and $\frac{dy}{db}$?

a. 1,0

b. 0,1

c. 0,0

d. 1,1

$y = \min(a,b); \; a > b$

$y = b$

$\frac{dy}{da} = 0; \quad \frac{dy}{db} = 1.$

Let's say vectors $\vec{a} = \{2; 4\}$ and $\vec{b} = \{n; 1\}$ forms the first two principle components after applying PCA. Under such circumstances, which among the following can be a possible value of n?

a. 2

b. -2

c. 0

d. 1

PC are orthogonal to each other $\Rightarrow \vec{a} \cdot \vec{b} = 0$

$\Rightarrow 2n + 4 = 0.$

$\Rightarrow n = -2$

$\Sigma \rightarrow$ Covariance

$\rightarrow$ symmetric matrix.

$\downarrow$ Eigen vector values.

orthogonal to each other.

$A = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ 2×1 $\quad \Sigma = \begin{bmatrix} a_1 a_1 & \\ & \end{bmatrix}$ 2×2

a_1 = gender

a_2 = profession.

A single hidden and no-bias autoencoder has 100 input neurons and 10 hidden neurons. What will be the number of parameters associated with this autoencoder?

a. 1000
b. 2000
c. 2110
d. 1010

$y = \min(a, b)$ and $a > b$. What is the value of $\frac{dy}{da}$ and $\frac{dy}{db}$?

a. 1, 0
b. 0, 1
c. 0, 0
d. 1, 1

Let's say vectors $\vec{a} = \{2; 4\}$ and $\vec{b} = \{n; 1\}$ forms the first two principle components after applying PCA. Under such circumstances, which among the following can be a possible value of n?

a. 2
b. -2
c. 0
d. 1

$A = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$

a_1 = gender

a_2 = profession.

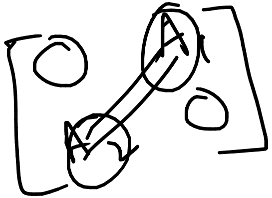
$$\Sigma_A = \begin{array}{c|cc} & a_1 & a_2 \\ \hline a_1 & \langle a_1, a_1 \rangle & \langle a_1, a_2 \rangle \\ a_2 & \langle a_2, a_1 \rangle & \langle a_2, a_2 \rangle \end{array}$$

univariate variance

statistical cross correlation.

univariate variance

corrcoef(A, B).



Consider the 2-layer neural network shown below. The weights are represented as follows: (w^k_{mn}) = weight between n^{th} node of k^{th} layer and m^{th} node $(k-1)^{th}$ layer. 0^{th} node is the bias node = 1 as depicted in the diagram.

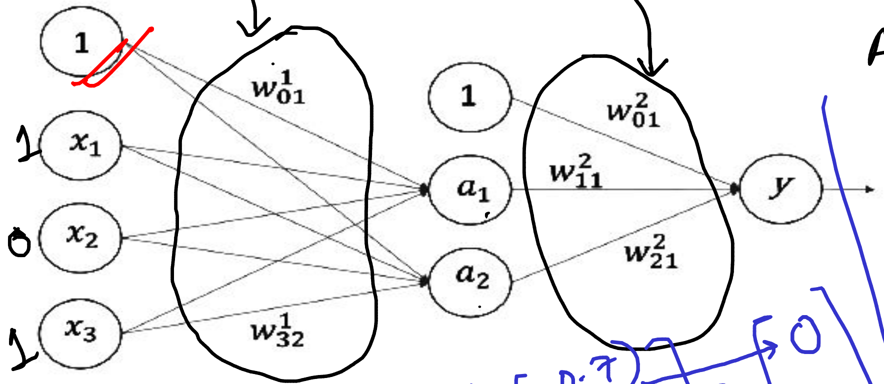
e.g. w^1_{32} = weight between 2^{nd} node of hidden layer and 3^{rd} node of input layer. Refer to the diagram. All weights have not been shown to maintain clarity.

Sigmoid activation function is applied to both the hidden layer and the output layer. The loss function is defined as $J(\cdot) = 0.5(y-t)^2$ where t is the true label.

The initial weights are given as:

$$W^1 = \begin{bmatrix} -0.4 & 0.2 & 0.4 & -0.5 \\ 0.2 & -0.3 & 0.1 & 0.2 \end{bmatrix} \qquad W^2 = [0.1 \quad -0.3 \quad -0.2]$$

Find the output at node a_1 and a_2 for given input $\{x_1 = 1, x_2 = 0, x_3 = 1\}$?



a. 0.13, 0.54
b. 0.33, 0.52
c. 0.23, 0.51
d. 0.13, 0.51

$$A = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}_{2\times1} \qquad X = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}_{4\times1}$$

$$A = \sigma(\underbrace{W^1 \cdot X}_{(2\times4)\,(4\times1)})$$

$$A = \sigma\left(\begin{bmatrix} -0.4 & 0.2 & 0.4 & -0.5 \\ 0.2 & -0.3 & 0.1 & 0.2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}\right)$$

$$= \sigma\left(\begin{bmatrix} -0.7 \\ 0.1 \end{bmatrix}\right) = \begin{bmatrix} \sigma(-0.7) \\ \sigma(0.1) \end{bmatrix} = \begin{bmatrix} 0.33 \\ 0.52 \end{bmatrix}$$

$$\begin{bmatrix} Relu(-0.7) \\ Relu(0.1) \end{bmatrix} = \begin{bmatrix} 0 \\ 0.1 \end{bmatrix}$$

Consider the 2-layer neural network shown below. The weights are represented as follows: w_{mn}^{k} = weight between n^{th} node of k^{th} layer and m^{th} node $(k-1)^{th}$ layer. 0^{th} node is the bias node $=1$ as depicted in the diagram.

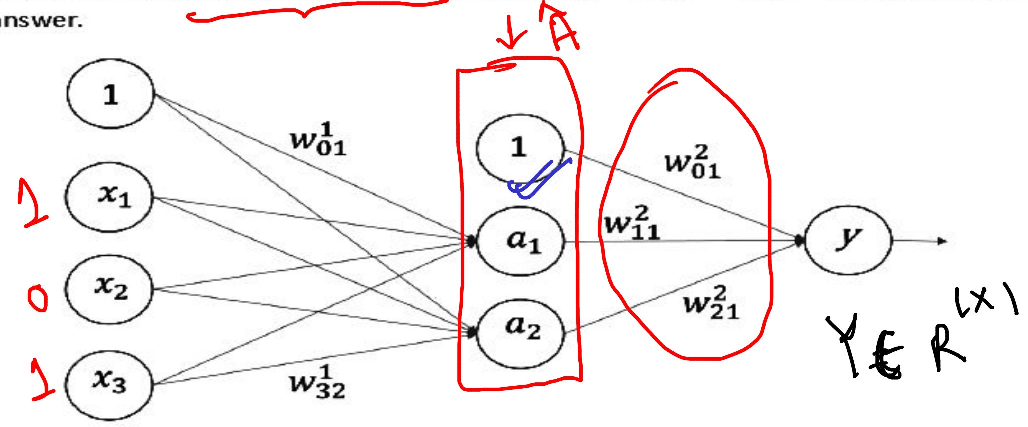
e.g. w_{32}^{1} = weight between 2^{nd} node of hidden layer and 3^{rd} node of input layer. Refer to the diagram. All weights have not been shown to maintain clarity.

Sigmoid activation function is applied to both the hidden layer and the output layer. The loss function is defined as $J(\cdot) = 0.5(y-t)^2$ where t is the true label.

The initial weights are given as:

$$W^1 = \begin{bmatrix} -0.4 & 0.2 & 0.4 & -0.5 \\ 0.2 & -0.3 & 0.1 & 0.2 \end{bmatrix} \quad W^2 = [0.1 \quad -0.3 \quad -0.2]$$

Find the final output at node y for given input $\{x_1 = 1, x_2 = 0, x_3 = 1\}$? Choose the closest answer.



a. 0.13
b. 0.33
c. 0.48
d. 0.51

$$\hat{A} = \begin{bmatrix} 1 \\ 0.33 \\ 0.52 \end{bmatrix}_{3\times1} \qquad W^2 = [0.1 \ -0.3 \ -0.2] \quad 1\times3$$

$$Y = \sigma(\underset{(1\times3)}{W^2} \cdot \underset{(3\times1)}{\hat{A}})$$

$$= \sigma\left(\underbrace{[0.1 \ \ -0.3 \ \ -0.2]\begin{bmatrix} 1 \\ 0.33 \\ 0.52 \end{bmatrix}}\right)$$

$$= \sigma(\underbrace{-0.103}) = \underline{\underline{0.48}}$$

Consider the 2-layer neural network shown below. The weights are represented as follows: w^k_{mn} = weight between n^{th} node of k^{th} layer and m^{th} node $(k-1)^{th}$ layer. 0^{th} node is the bias node = 1 as depicted in the diagram.

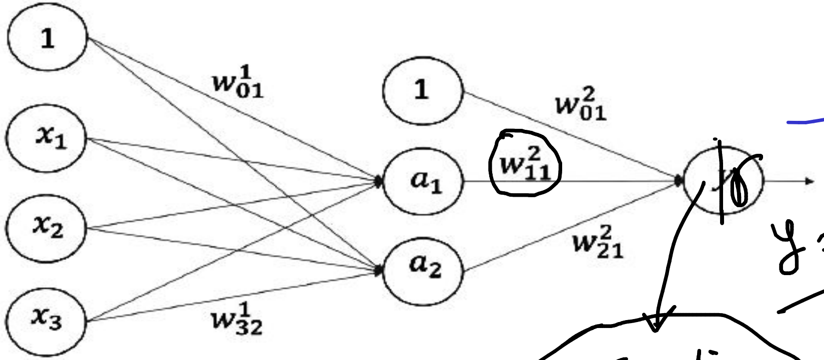
e.g. w^1_{32} = weight between 2^{nd} node of hidden layer and 3^{rd} node of input layer. Refer to the diagram. All weights have not been shown to maintain clarity.

Sigmoid activation function is applied to both the hidden layer and the output layer. The loss function is defined as $J(\cdot) = 0.5(y - t)^2$ where t is the true label.

The initial weights are given as:

$$W^1 = \begin{bmatrix} -0.4 & 0.2 & 0.4 & -0.5 \\ 0.2 & -0.3 & 0.1 & 0.2 \end{bmatrix} \quad W^2 = [0.1 \quad -0.3 \quad -0.2]$$

Find the gradient component $\frac{\partial J}{\partial w^2_{11}}$ for $t = 1$ and given input $\{x_1 = 1, x_2 = 0, x_3 = 1\}$? Choose the closest answer.



a. -0.09
b. -0.11
c. -0.13
d. -0.04

$$\frac{\partial J}{\partial W^2_{11}}$$

$$f(x) = x^2 + x^3$$

$$J(y) = \frac{1}{2}(y-t)^2 \; ; \; P = W^2_{01} + W^2_{11}\cdot a_1 + W^2_{21}\cdot a_2 .$$

$$y = \sigma(P)$$

$\frac{\partial y}{\partial P}$ = derevative of sigmoid function.

$$= \sigma(P)\cdot(1-\sigma(P))$$

$$= y\cdot(1-y)$$

$$\frac{\partial J}{\partial W^2_{11}} = \frac{\partial J}{\partial y} \frac{\partial y}{\partial P} \frac{\partial P}{\partial W^2_{11}}$$

$$y = \sigma(P)$$

$$P = \sum W_i x_i$$

$$\frac{\partial J}{\partial W^2_{11}} = (y-t)\cdot y(1-y)\,a_1$$

$$= (0.48-1)\,0.48\,(1-0.48)\,0.33$$

$$= -0.04.$$

Consider the 2-layer neural network shown below. The weights are represented as follows: w_{mn}^{k} = weight between n^{th} node of k^{th} layer and m^{th} node $(k-1)^{th}$ layer. 0^{th} node is the bias node = 1 as depicted in the diagram.

$$f = \max(0, x)$$

e.g. w_{32}^{1} = weight between 2^{nd} node of hidden layer and 3^{rd} node of input layer. Refer to the diagram. All weights have not been shown to maintain clarity.

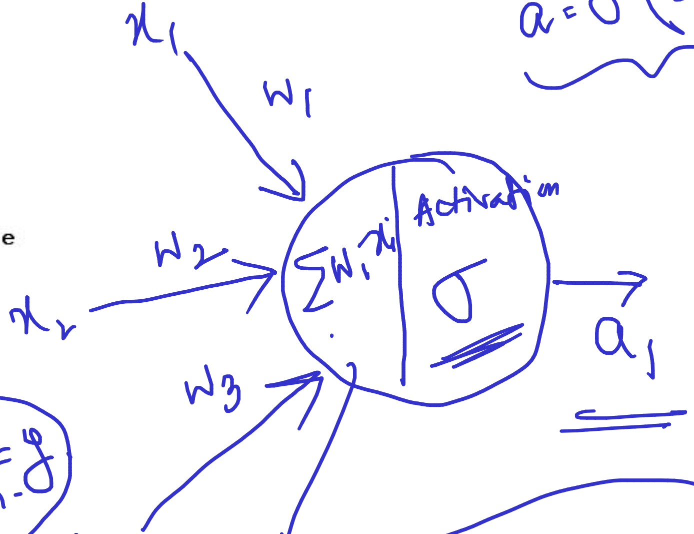
Sigmoid activation function is applied to both the hidden layer and the output layer. The loss function is defined as $J(\cdot) = 0.5(y-t)^2$ where t is the true label.

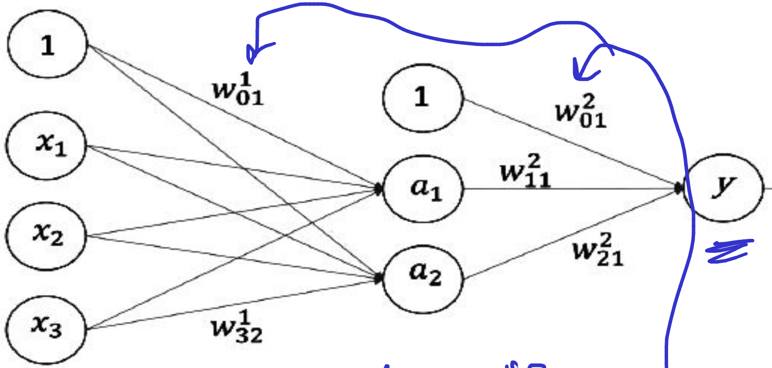
$$a = \sigma\left(\sum w_i x_i\right)$$

The initial weights are given as:

$$W^1 = \begin{bmatrix} -0.4 & 0.2 & 0.4 & -0.5 \\ 0.2 & -0.3 & 0.1 & 0.2 \end{bmatrix} \quad W^2 = [0.1 \quad -0.3 \quad -0.2]$$

Find the gradient component $\frac{\partial J}{\partial w_{11}^{2}}$ for $t = 1$ and given input $\{x_1 = 1, x_2 = 0, x_3 = 1\}$? Choose the closest answer.





Model prediction $= \hat{y}$

$L = \frac{1}{2}(0.48 - 1)^2$

$w_1x_1 + w_2x_2 + w_3x_3 = \sum_i w_i x_i$

$y = 0.48.$

$t = 1.$

$L =$

a. −0.09
b. −0.11
c. −0.13
d. −0.04

Consider the 2-layer neural network shown below. The weights are represented as follows: w^k_{mn} = weight between n^{th} node of k^{th} layer and m^{th} node $(k-1)^{th}$ layer. 0^{th} node is the bias node = 1 as depicted in the diagram.

e.g. w^1_{32} = weight between 2^{nd} node of hidden layer and 3^{rd} node of input layer. Refer to the diagram. All weights have not been shown to maintain clarity.

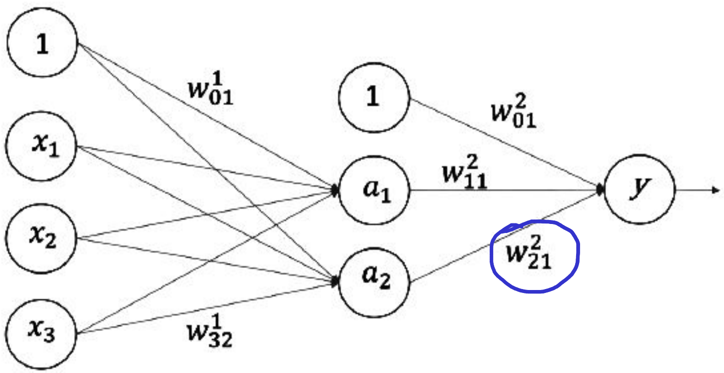
Sigmoid activation function is applied to both the hidden layer and the output layer. The loss function is defined as $J(\cdot) = 0.5(y-t)^2$ where t is the true label.

The initial weights are given as:

$$W^1 = \begin{bmatrix} -0.4 & 0.2 & 0.4 & -0.5 \\ 0.2 & -0.3 & 0.1 & 0.2 \end{bmatrix} \quad W^2 = [0.1 \ \ -0.3 \ \ -0.2]$$

$W^2_{21} = -0.2$

Find the updated value of w^2_{21} after 1 iteration for $t = 1$, the learning rate $\eta = 0.9$ and given input $\{x_1 = 1, x_2 = 0, x_3 = 1\}$? Choose the closest answer.



a. -0.29
b. -0.1
c. -0.14
d. -0.04

W^2_{21} after 1^{st} iteration :→

$$\underbrace{W^{2(1)}_{21}}_{1^{st}\text{ iteration}} \leftarrow W^{2(0)}_{21} - \eta \frac{\partial J}{\partial W^2_{21}}$$

$$\frac{\partial J}{\partial W^2_{21}} = \frac{\partial J}{\partial y} \frac{\partial y}{\partial p} \frac{\partial p}{\partial W^2_{21}}$$

$$= (y-t)\, y(1-y) \cdot a_2$$

$$\frac{\partial J}{\partial W^2_{21}} = -0.067$$

$$W^{2(1)}_{21} \leftarrow (-0.2) - [0.9 \times -0.067]$$

$$= -0.139$$

$$\approx -0.14$$

Select the correct option.

Initializers → He, Glorot, Uniform.

$W(n) \leftarrow W(n-1) - \eta \frac{dE}{dW}$

a. Layer-by-layer autoencoder pretraining reduces GPU/CPU RAM requirements

b. Layer-by-layer autoencoder pretraining alleviates slow convergence

c. Layer-by-layer autoencoder pretraining followed by finetuning converges to more optimal parameters than End-to-End training of autoencoders

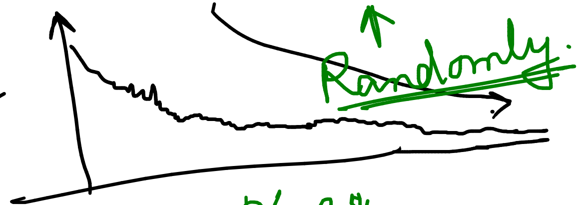
d. All of the above

10 / 1m

$n=1$

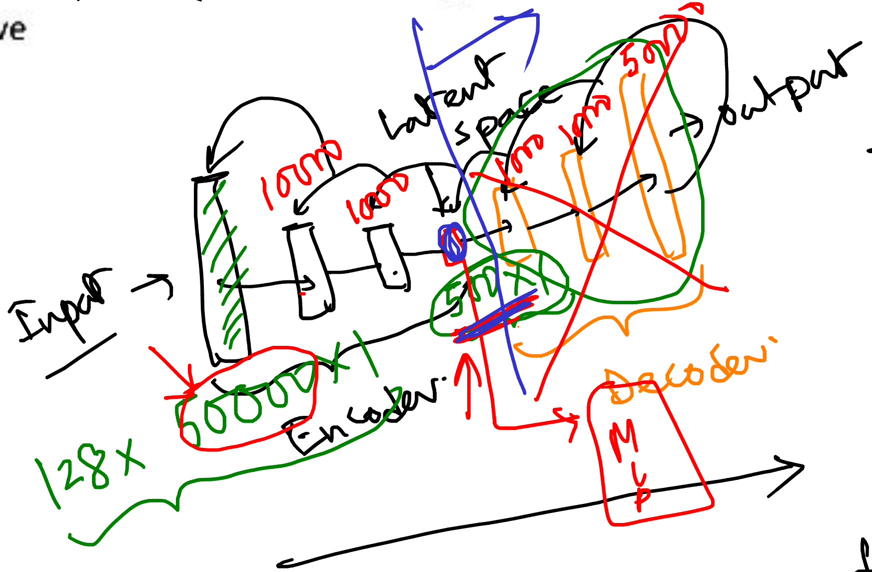
$W(1) \leftarrow W(0) - \eta \frac{dE}{dW}$

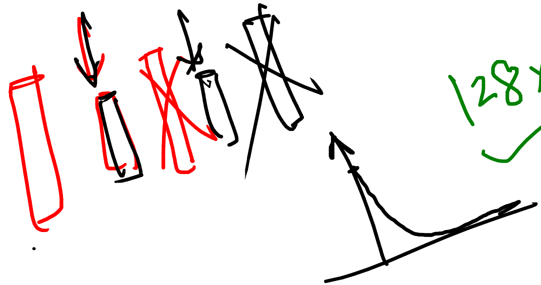
Randomly.



2% - 3%

Stacked AE





If requires high RAM GPU memory

Training of SAE from scratch then Too many parameter has to be optimized

Regularization of Contractive Autoencoder is imposed on

a. Jacobian matrix of encoder activations with respect to the input
b. Weights
c. Inputs
d. Does not use regularization

$$\begin{bmatrix} \frac{\partial A(x)}{\partial x_1} & \frac{\partial A(x)}{\partial x_2} & \cdots \\ & & \end{bmatrix}$$

Select true statements about KL Divergence

(incorrect)

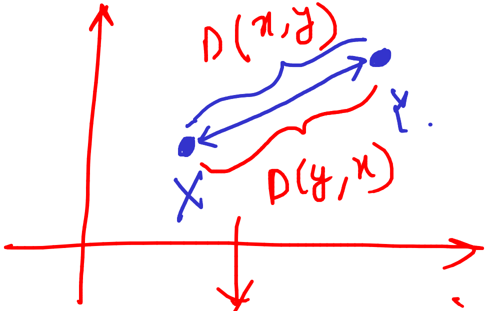
incorrect

a. Measures ~~distance~~ between two probability distribution

b. Has range from 0 to 1 $0 \to \infty$

c. Is symmetric, i.e. $KL(P|Q) = KL(Q|P)$ (incorrect)

d. None of above



Distance is symmetric in nature

$$KL\ divergen\ (P|Q) = P \log \frac{P}{Q} = P \log P - P \log Q.$$

$$KL\ div\ (Q|P) = Q \log \frac{Q}{P} = Q \log Q - Q \log P.$$

$$KL(P|Q) \neq KL(Q|P).$$

$P, Q = 0.$ $P \cdot \log \frac{P}{0}$ $P \log(\infty)$ $P \times \infty = \infty$

$P = 0,\ \left(P \log \frac{P}{Q}\right) = 0 \cdot \left(\log \frac{0}{Q}\right)$ → undefined

$P = 0,\ Q = 0$ $0 \cdot \log\left(\frac{0}{0}\right) = 0.$

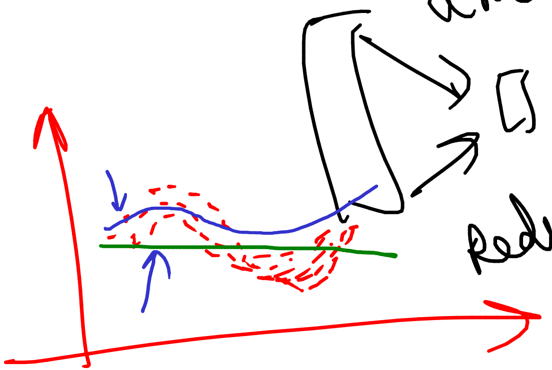
Range $0 \to \infty$

In which conditions, autoencoder has more powerful generalization than Principal Components Analysis (PCA) while performing dimensionality reduction?

linear transform.

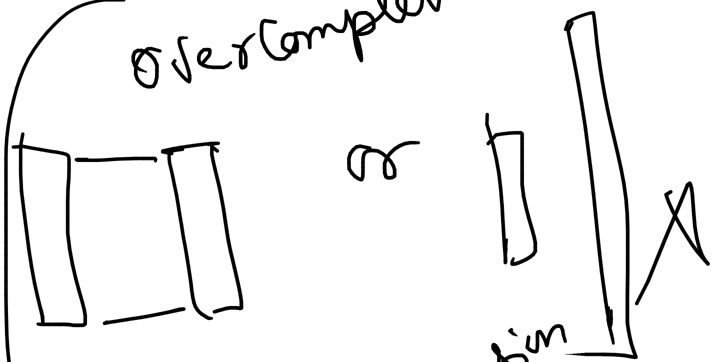
a. Undercomplete Linear Autoencoder
b. Overcomplete Linear Autoencoder
c. Undercomplete Non-linear Autoencoder
d. Overcomplete Non-linear Autoencoder

under complete.



Reduces dimen

Overcomplete



either same dimension or increased dimension to that of input size

An autoencoder consists of 128 input neurons, 32 hidden neurons. If the network weights are represented using single precision floating point numbers (size= 4 bytes) then what will be size of weight matrix?

a. $33408\ Bytes$
b. $16704\ Bytes$
c. $8352\ Bytes$
d. $32768\ Bytes$

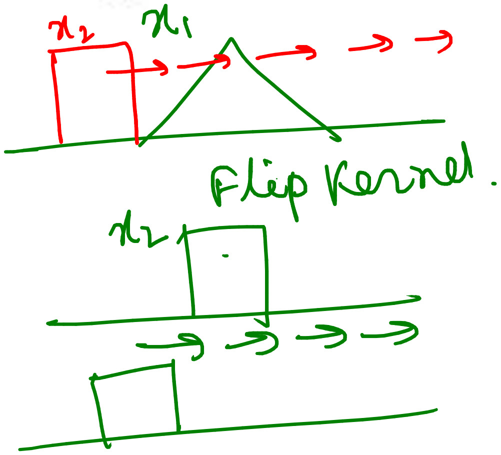
Which of the following is used to match template pattern in a signal

a. Cross Correlation

b. Convolution

c. Normalized cross correlation

d. None of the above



Lecture 35 at 28:49

Which of the following is an LTI/LSI system? y and x are output and input respectively.

→ Super position theorem

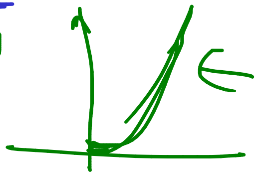
a. $y = m \times x + n \times x$

b. $y = m \times x + c$

c. $y = m \times x - c$

d. $y = m \times x^2$

Non linear



← Non linear

(a) $y = (m+n)x$.

Superposition theorem obeyed

$x_1 :- \; y_1 = (m+n)x_1 = mx_1 + nx_1$

$x_2 :- \; y_2 = (m+n)x_2 = mx_2 + nx_2$

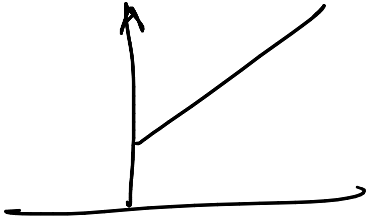
Now, $y_F = (m+n)(x_1+x_2)$

$= mx_1 + mx_2 + nx_1 + nx_2$

$= mx_1 + nx_1 + mx_2 + nx_2$

$y_F = y_1 + y_2$

(b) $y = mx + C$.



x_1 , x_2

$y_1 = mx_1 + C$.

$y_2 = mx_2 + C$.

$y_1 + y_2 = mx_1 + mx_2 + 2C$

you apply $(x_1 + x_2)$

$y_3 = m(x_1 + x_2) + C$.

$y_3 \neq y_1 + y_2$ → super position theorem is not valid

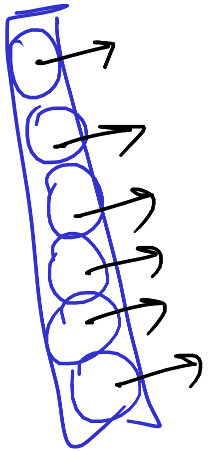
What is the role of sparsity constraint in a sparse autoencoder?

a. Control the number of active nodes in a hidden layer

b. ~~Control the noise level in a hidden layer~~

c. ~~Control the hidden layer length~~

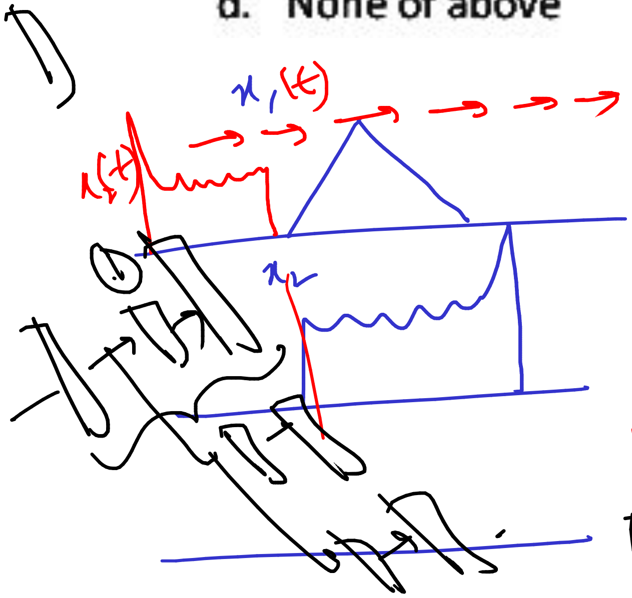
d. ~~Not related to sparse autoencoder~~

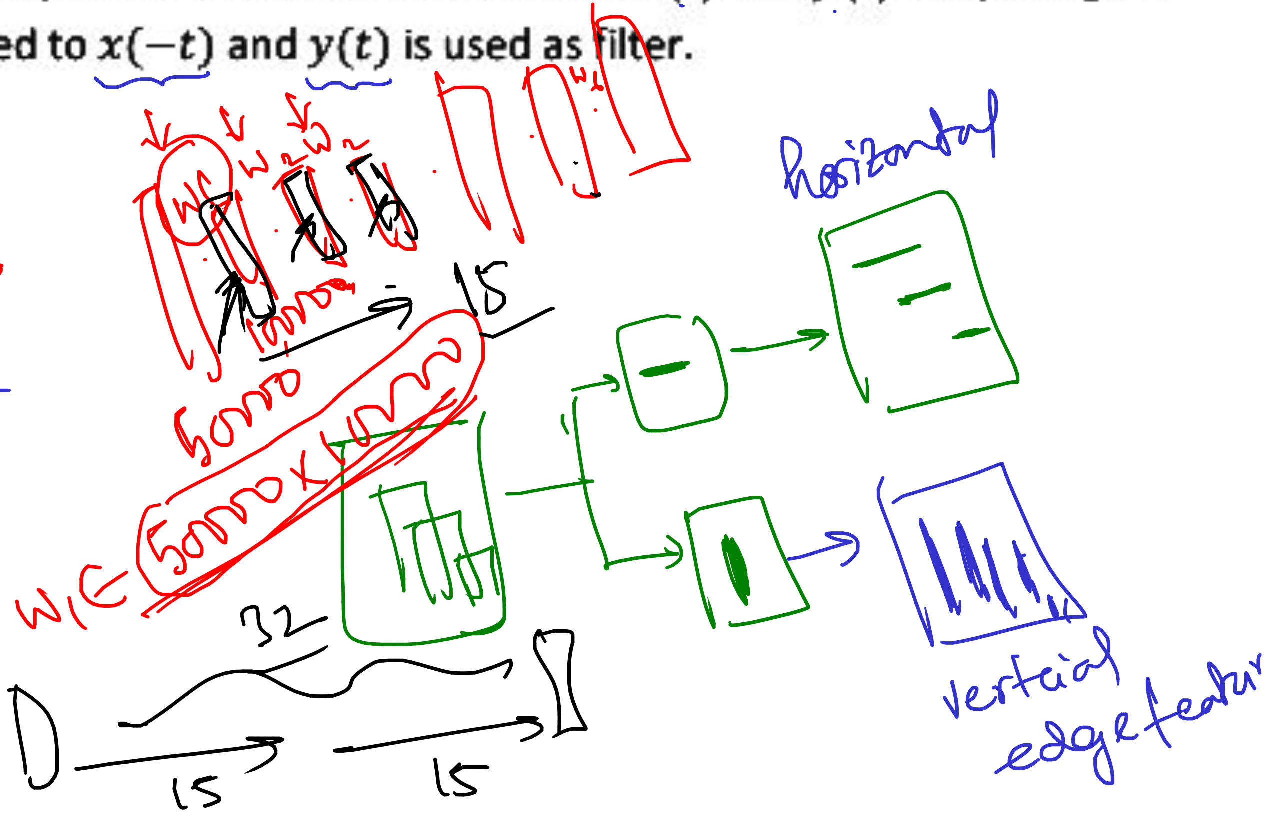


Aim to control the activation of the neurons in the hidden layer

Which of the following is true about convolution?

a. Convolution is used to compute features from signal
b. Can be used to compute cross correlation between $x(t)$ and $y(t)$ if input signal $x(t)$ is transformed to $x(-t)$ and $y(t)$ is used as filter.
c. Both a and b
d. None of above





Thank you