classmate
Date
Page

Unit 4 : AutoEncoders.

Autoencoders.

1) Undercomplete Autoencoders.
2) Regularized Autoencoders.
3) Representational power, Layer size and Depth
4) Stochastic Encoders & Decoders :
5) Denoising Autoencoders.
6) Contractive Autoencoders.
7) Applications of Autoencoders.
8) Variational Autoencoders.

Questions from previous CIE :

Q) Exemplify Under Complete Autoencoders, Sparse AutoEncoders, and Stochastic AutoEncoders in detail (Layer, size, depth, & basic concepts with equations) (10 marks).

Q)

Questions from previous SEE :

Q) What are undercomplete Autoencoders? Identify the significance of them (10m).

Q) Discuss the various applications of auto encoders (10m).

(or)

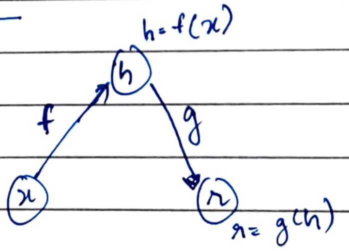
Q) Investigate how to model an AutoEncoder that is capable of performing even when the i/ps are corrupted (or) with noise (10m).

Q) Illustrate Stochastic Encoders & decoders with relevant diagram & equations.

Auto Encoders.

- An autoencoder is a Neural Network that is trained to attempt to copy its input to its output.
- Internally it has a hidden layer 'h' that describes a code used to represent the input (i.e encode the input)
- An autoencoder n/w consists of 2 parts :
  - i) an encoder function : $h = f(x)$
  - ii) a decoder function that tries to reconstructs the input : $\hat{x} = r = g(h)$

- If an autoencoder learns to successfully set $g(f(x)) = x$ everywhere, then it is not useful.
  Instead, they are restricted to only copy approximately, & to copy only the input that resembles the training data. Because the model is forced to learn only the useful aspects of the data.



- Autoencoders are used for dimensionality reduction or feature learning.

1) Undercomplete Autoencoders.

- An autoencoder whose encoded dimension (i.e dimension of h) is less than the input dimension (i.e dim x) is called an undercomplete autoencoder.
- Learning an effective undercomplete representation forces the autoencoder to capture the most salient features of the training data.

- The learning process for an autoencoder is described as minimizing a loss function: $L(x, g(f(x)))$

where L is a loss function penalizing $g(f(x))$ to be dissimilar from x. (eg: MSE).

- when the decoder (g) is linear & L is the mean squared error, an Undercomplete autoencoder learns to span the same subspace as PCA.
- Autoencoders with non-linear encoder functions (f) and non-linear decoder functions (g) can learn a more powerful non linear generalization of PCA.

2) Regularized Autoencoders.

- Regularized autoencoders provide the ability to train any architecture of the autoencoder successfully i.e choosing the code dimension (i.e $\dim^{n}$ of h), & the capacity of the encoder & decoder based on the complexity of the distribution to be modeled.
- Regularized autoencoders allows to do so by using a loss function that encourages the model to have other properties besides the ability to copy its input. These other properties include
  - i) sparsity of the representation.
  - ii) smallness of the derivative of the representation.
  - iii) robustness to noise or missing inputs.

2.1) Sparse Autoencoders.

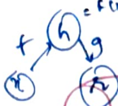
- A sparse autoencoder is simply an autoencoder whose training criterion includes a sparsity penalty $\Omega(h)$ on the encode layer h, in addition to the reconstruction error

i.e $L(x, g(f(x))) + \Omega(h)$

where $h = f(x)$ is the encoder o/p

& $g(h)$ is the decoder o/p.

- Sparse autoencoders are typically used to learn features



for another task, such as classification.

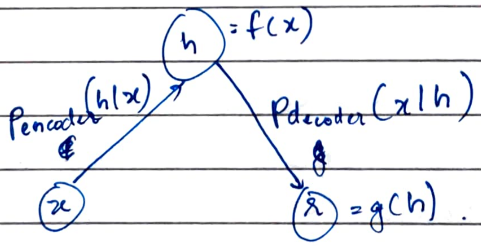
:) A sparse autoencoder responds only to the unique statistical features of the training dataset, rather than simply acting as an identity function.

4) Stochastic Encoders & Decoders.

.) Autoencoders are just FFNN. But in an autoencoder, the input x is also the target.

:) Given a hidden code $h = f(x)$, we may think of the decoder as providing a conditional distribution $P_{decoder}(x|h)$.

⇒ We may then train the autoencoder by minimizing $-\log P_{decoder}(x|h)$.



~~[illegible]~~ in general

stochastic encoder: $P_{encoder}(h|x) = P_{model}(h|x)$.

stochastic decoder: $P_{decoder}(x|h) = P_{model}(x|h)$.

5) Denoising Autoencoders (DAE)

.) A denoising autoencoder is an autoencoder that receives a corrupted datapoint as input & is trained to predict the original, uncorrupted datapoint as its output.

.) Traditionally autoencoders minimize the loss function of the form

$$L(x, g(f(x)))$$

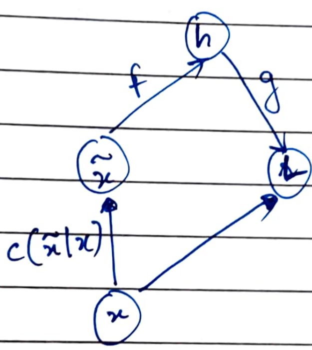
where L is the loss function ~~that~~ penalizing $g(f(x))$ being ~~dis~~ dissimilar from x.

- A DAE instead minimizes ~~the~~ a loss function of the form.

$$L(x, g(f(\tilde{x}))).$$

where $\tilde{x}$ is a ~~copy~~ ~~of~~ the input that has been corrupted by some form of noise & the output is the original & uncorrupted 'x'.

- The DAE training process is illustrated in the figure below



- We introduce a corruption process $C(\tilde{x}|x)$, which represents the conditional distribution of the corrupted samples $\tilde{x}$, given a data sample x.

- The autoencoder then learns a reconstruction distribution $P_{reconstruct}(x|\tilde{x})$ from the training pairs $(x, \tilde{x})$ as follows

  i) Sample a training example x ~~fr.~~ from the training data.

  ii) Sample a corrupted version ~~a~~ $\tilde{x}$ from $C(\tilde{x}|x)$.

  iii) Use $(x, \tilde{x})$ as ~~the~~ a training example for estimating the autoencoder reconstruction distribution

  ~~BC~~ $P_{reconstruct}(x|\tilde{x}) = P_{decoder}(x|h)$ ~~with~~

  with h = output of the encoder $f(\tilde{x})$

7) Applications of Autoencoders.

·) 2 main applications of Autoencoders
i) Dimensionality Reduction
ii) Information Retrieval.

i) Dimensionality Reduction
·) Lower dimensional representations can improve the performance of many tasks, such as classification.
·) Models of smaller spaces consume less memory & runtime.
·) It also helps ~~gear~~ in better model generalization.

ii) Information Retrieval.
·) Semantic hashing.