

# ACVDL Suggestions

ATTEMPT ALL (No alternatives)  
ALL Question  $\geq$  5 marks  
Transformer Question Guaranteed

Written by Sayan

Das

## Lec3 PDF

- (pg 4) Derivation of geometric margin
- (pg 6)  $\gamma_{new}$  greater than equal to  $\gamma_{old}$
- (pg 7) Algorithm (exactly as the notes)
- (pg 10) lowest no of nodes and connections that classifies xor and their weights and thresholds

## Lec4 PDF

- (pg 9) Softmax and its gradient derivation
- (pg 12) Binomial to Binary cross entropy loss and Categorical Distribution to Categorical Cross entropy loss
- (pg 15) KL Divergence: why is it not a distance metric and how to make it symmetric? And whether it is still a distance metric?

## Lec5 PDF

- (pg 2) Show that combination of multiple linear neurons still make it linear (for n layers). [Gemini Link](#)
- (pg xx) Given a function prove whether the given function is convex or not. Use Hessian Method. This can be two layered neural network with two nodes in each layer and a sigmoid activation at the final layer. [Gemini Link](#)

## Lec6 PDF

- Calculate the layer-wise parameter and the full parameter along with the feature volume of each of the output layers of a particular neural network with the following layers:
  - Conv2D
  - Conv1D
  - MaxPool(1D or 2D)
  - BatchNorm(1D or 2D)
  - Linear (Fully Connected)
  - Any activation

Any dropout  
Upsampling  
Conv2DTranspose  
Dilated Convolution (15-20 marks)

## Lec7 PDF

- (pg 1) Show that any linear layer neural network with multiple layers and encoder decoder architecture converges to a linear PCA. Hence, derive principal components for a given  $n$  data points in  $R^d$ . [Gemini Link Good](#)
- (pg 4) Is the posterior distribution in VAEs intractable? and why? [Gemini Link](#)
- (pg 9) Derive the Evidence Lower Bound or Variational Lower Bound.
- (pg 13) What is the problem with standard sampling in VAEs and how is reparametrisation trick used in VAEs and what is reparametrisation trick?

## Lec8 PDF

- (pg 5) How is CBOW different from skip-gram model. Draw and Explain. Also write how can they be used to learn word vectors from text corpus.
- (pg 8) For a simple RNN write the backpropagation through time (BPTT) for all the layers [Gemini Link](#)

## Lec9 PDF (check notes)

- (pg 4) Draw a LSTM Network and write all of its components i.e. all the gates and all the layers mathematically and explain their functionalities. [Gemini Link](#)
- Draw a GRU Network and write all of its components i.e. all the gates and all the layers mathematically and explain their functionalities. [Gemini Link](#)
- (pg 5) Write all the 5 types of attention mathematically and write about each of the components.

## Lec10-11 PDF

- (lec 10, pg 11) Draw the transformer network with all the bells and whistles. Draw the encoder, decoder, positional embedding, attention, Add and Norm, query-key-value, fully connected layer, multi-head attention, cross attention, layer normalization, the mathematics for positional embedding and the mathematics for KQV attention. \*\*\*\* (lec 11, pg 5) Transformer Architecture Diagram Show how the word is being added with positional encoding and all the other steps involved during the contextualization of the word. (30 mark)
- (lec 11, pg10) What does a multi-head attention do? show with matrix diagram.