|  |  |  |
| --- | --- | --- |
| UNIT-1 | | |
|  | Write the equations of Norm Penalties | 2 |
|  | 1. What is a learning rate, and how is it useful in weight updates? 2. What is a vanishing gradient, and how does it effect the back-propagation algorithm? 3. What is the kernel used for vertical feature identification? 4. What will happen if the kernel size is increased or decreased? 5. What is a ‘Grandmother neuron'? |  |
|  | How do you decide which activation function to use in a neural network? |  |
|  | Given a function y = (x + 4)2. Evaluate the local minima of the function starting from the point x = 3 and the value of x after the first iteration using gradient descent (Assume the learning rate is 0.01)? |  |
|  | Represent sigmoidal activation functions with figure and equation. |  |
|  | Explain the effect of ‘’ in learning stage of the model. Mention the key issue of choosing high and low learning rate. |  |
|  | Analyse the NN cost function |  |
|  | 1. In a convolution neural network, how is a kernel used, and mention the standard kernel used to identify the vertical features? 2. Write the standard size of the kernel and Justify the usefulness of it. 3. Does every human has a ‘Halle Berry neuron’? 4. Draw a diagram representing the local minimum, local maximum, global minimum, and global maximum. 5. What are the different types of activation functions used in Multilayer perceptrons? |  |
|  | What impact does the output unit have on model performance |  |
|  | Differentiate overfitting and under fitting |  |
|  | What is regularization how does it make deep learning model better |  |
|  | Define Gradient descent |  |
|  | 1. Write the weight update rule equation and explain each term of the equation. 2. What is stride, and how does it help to apply a kernel? 3. What is a kernel in CNN, and what are the standard kernels used? 4. What is a ‘Halle Berry neuron’? 5. Draw a diagram representing the local minimum, local maximum, global minimum, and global maximum. |  |
|  | ESSAY: |  |
|  | 1. Explain Gradient descent in the context of deep learning and with a neat diagram representing plateau and saddle points. 2. Explain the concept of perceptron with a neat diagram. And also explain why it is not suitable for non-linear datasets. |  |
|  | Apply backpropagation on the following NN with sigmoid activation function, learning rate=1, actual output=1 calculate error for one forward pass, the updated weights and bias for one backward pass  Take a sample network | 5 |
|  | Give clear descriptive reasons on why regularization with constraints is chosen than by norm penalties. |  |
|  | a) Explain how the XOR problem will be solved using deep neural network (3M)  b)Describe Regularization techniques in deep neural networks(2M) |  |
|  | Apply backpropagation on the following NN with sigmoid activation function, learning rate=0.8, actual output=1 and MSE loss function. Perform forward pass, backward pass and another forward pass and backward pass  Take a sample netwrk |  |
|  |  |  |
|  | Explain the types of Gradient Decent methods |  |
|  | Illustrate Data Augmentation and its application in training the model. Interpret bias and variance tradeoff. |  |
|  | Justify with necessary derivation that momentum based gradient descent reaches faster to the convergence point |  |
|  | Compare Feed Forward and Back Propagation Neural Network. Show BPN is a supervised learning. |  |
|  | Describe gradient descent and its types. Establish the weight adjustment strategy for the NN model's training. |  |
|  | Explain about learning conditional statistics in gradient based learning |  |
|  | Why does a single perceptron cannot simulate simple XOR function? Explain how to overcome this. |  |
|  | Design a deep feedforward network for a classification problem and explain how do the different layers work in the network. |  |
|  | Explain ridge and lasso regression with equations |  |
|  | Explain Backpropagation algorithm and its working with own network |  |
|  | Construct a convolutional network to demonstrate the use of Zero padding on an input image (take your own image size kernel and filters) |  |
|  | How do you speed up convolutional algorithms . explain briefly in terms of time complexity and operations. |  |
|  | Compare L1 and L2 regularizations with respect to test error and train error |  |
|  |  |  |
|  |  |  |
| UNIT-2 | |  |
|  | List any two types of Data used by a convolutional network | 2 |
|  | Outline the strategies to obtain convolution kernels without supervised learning | 2 |
|  | Evaluate the size of a feature map, given that the image size is 32x32, filter size is 5x5, stride is 1, and no padding |  |
|  | State different layers used in CNN. |  |
|  | Differentiate between NN and CNN. Show CNN is powerful than NN. |  |
|  | Compare filter and feature map of CNN |  |
|  | Show convolution operation on a5X5 image with a kernel of size 3X3 with a stride of 2. What is size of output |  |
|  | What are the three stages of convolution network |  |
|  | An input with image size 28X28 ,filter size 3X3, padding 2, stride 2.what is output size of the image |  |
|  | What is separable convolution |  |
|  |  |  |
|  |  |  |
|  | ESSAY: |  |
|  | Which three primary Visual cortex V1 properties are loosely used to design a Convolutional Neural Network? |  |
|  | Perform the convolutional operations on the given input image color values and the given kernel. Note stride = 1 and also apply max pool of size 2\*2   |  |  |  |  |  | | --- | --- | --- | --- | --- | | 168 | 168 | 184 | 191 | 165 | | 178 | 251 | 197 | 100 | 137 | | 184 | 181 | 161 | 190 | 174 | | 191 | 197 | 186 | 175 | 135 | | 92 | 165 | 139 | 173 | 187 |  |  |  |  | | --- | --- | --- | | -1 | 1 | 0 | | 1 | -1 | -1 | | 1 | -1 | 0 |   input image color values  kernel |  |
|  | With a neat diagram explain various architectural key components of CNN? | 5 |
|  | Given 13 pixel values (1,2), (2,2) (3,1) (3,4) (3,7) (4,2) (4,1) (4,4) (4,5) (5,1) (5,5) (6,2) (6,4) for example, apply 3-Means algorithm and show why unsupervised algorithms are preferred in convolution filters? |  |
|  | convolution layer and Pooling layers in CNNs? (1M)  b) Apply convolution on 4×4 input [[8,5,-3,7],[2,4,1,-6],[3,-4,-9,7],  [5,-1,6,0]] with 2\*2 filter [[5, 7], [-2,3]] with a) stride 1 and stride 2. b) apply convolution with padding c) Apply max pooling on the above result (4M) |  |
|  | 1. Perform the convolutional operations on the given input image color values and the given kernel. Note stride = 1 also apply max pool of size 2\*2  |  |  |  |  |  | | --- | --- | --- | --- | --- | | 168 | 188 | 164 | 161 | 195 | | 178 | 201 | 197 | 150 | 137 | | 174 | 161 | 181 | 190 | 184 | | 131 | 179 | 176 | 185 | 195 | | 92 | 185 | 179 | 133 | 167 |  |  |  |  | | --- | --- | --- | | 0 | 1 | 0 | | 1 | 1 | 1 | | 0 | 1 | 0 |   kernel  input image color values |  |
|  | a) Explain about variants of basic convolution function. (2M)  b) Describe neuro scientific basis for Convolution networks (3M) |  |
|  | Illustrate the effect of zero padding on convolution network with a kernel of width six at every layer with pictorial representation |  |
|  | 1. What is regularization? And explain different types of regularization techniques. 2. Give the basic architecture of Convolution neural networks and explain each layer. |  |
|  | Explain the working of convolution neural network |  |
|  | Derive structure of the below image after flattening.   |  |  |  | | --- | --- | --- | | 1 | 2 | 2 | | 2 | 5 | 2 | | 2 | 7 | 1 |   Derive output of below image after passing through Relu activation function.   |  |  |  |  | | --- | --- | --- | --- | | -2 | 2 | 0 | -3 | | -2 | 3 | -2 | -1 | | 1 | 5 | 6 | 0 | | -4 | 5 | -4 | 0 | |  |
|  | Build the structure of flatten layer when below 4 X 4 image passes through 3 X 3 kernel and 2 X 2 average pooling layer. Assume that Stride S=1, and Padding P=0.   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | 1 | 2 | 2 | 1 |  | 1 | 0 | 1 | | 1 | 2 | 2 | 1 |  | 1 | 0 | 1 | | 1 | 2 | 2 | 1 |  | 1 | 0 | 1 | | 1 | 2 | 2 | 1 |  | Kernel Size | | | | | Image Size | | | |  | |  |
|  | Build the convolutional image by applying the 3 × 3 kernel on to 6 × 6 grayscale image below. Consider Padding P=0 and Stride S=1.     |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 0 | 1 | 0 | 1 | 0 | 1 |  |  |  |  | | 0 | 1 | 0 | 1 | 0 | 1 |  | -1 | 1 | 1 | | 0 | 1 | 0 | 1 | 0 | 1 |  | -2 | 1 | 2 | | 0 | 1 | 0 | 1 | 0 | 1 |  | -1 | 1 | 1 | | 0 | 1 | 0 | 1 | 0 | 1 |  | Kernel (3 × 3) | | | | 0 | 1 | 0 | 1 | 0 | 1 |  |  |  |  | | image (6 × 6) | | | | | |  |  |  |  | |  |
|  | Identify the size of the CNN when kernel, padding and strides are known.  Determine the image's outputs, after applying the max, average, and sum pooling layers to the image below.   |  |  |  |  | | --- | --- | --- | --- | | 1 | 2 | 4 | 2 | | 6 | 6 | 5 | 5 | | 2 | 2 | 1 | 5 | | 6 | 2 | 2 | 4 | |  |
|  | What is pooling? Write on various pooling techniques. |  |
|  | Write on Neuro scientific basis for convolutional Networks. |  |
|  | Write on the variants of convolution function |  |
|  | Compare and contrast the operations of primary visual cortex with CNN. |  |
|  | 1. Give the basic architecture of Convolution neural networks and explain each layer. 2. Perform the convolutional operations on the given input image color values and the given kernel. Note stride = 1 and also apply max pool of size 2\*2  |  |  |  |  |  | | --- | --- | --- | --- | --- | | 168 | 188 | 164 | 161 | 195 | | 178 | 201 | 197 | 150 | 137 | | 174 | 161 | 181 | 190 | 184 | | 131 | 179 | 176 | 185 | 195 | | 92 | 185 | 179 | 133 | 167 |   kernel   |  |  |  | | --- | --- | --- | | -1 | 1 | 0 | | 1 | -1 | -1 | | 1 | -1 | 0 |   input image color values |  |
|  | Write on local connections, convolution and full connections with neat sketch. |  |
|  | How can convolutional kernels be learned without supervised learning models? Explain in detail. |  |
|  | Approaches of obtaining convolutional kernels without supervised learning |  |
|  |  |  |
|  |  |  |
| UNIT-3 | | |
|  | Compare Recurrent Neural Networks and Bidirectional RNNs | 2 |
|  | Define sequential modelling. What makes RNNs suitable for sequential data? | 2 |
|  | Explain the role of forget gate in an Long Short-Term Memory cell |  |
|  | Write the difference between LSTM and RNN |  |
|  | Illustrate different gates used in LSTM |  |
|  | How do Gated RNNs help recurrent NN learn long term dependencies. |  |
|  | Compare FFNN and RNN |  |
|  | Write on GRU and give its architecture |  |
|  | What is sequence modelling |  |
|  | Discuss the relevance LSTM over RNN |  |
|  |  |  |
|  | ESSAY: |  |
|  | What is sequence data. Which model is preferred to train on sequence data. Explain in brief. | 5 |
|  | Explain the fundamentals of Recurrent Neural Networks (RNNs) and their architecture for processing sequential data. Discuss the concept of recurrent connections and how RNNs maintain memory across time steps. |  |
|  | Analyze the architecture of Gated Recurrent Units (GRUs) and their mechanisms for handling the vanishing gradient problem in Recurrent Neural Networks (RNNs). Discuss the differences between GRUs and other recurrent units, such as LSTMs, and their relative effectiveness in different tasks. |  |
|  | Evaluate the challenges of training Recurrent Neural Networks (RNNs), such as vanishing and exploding gradients. Discuss techniques to address these challenges, such as gradient clipping and Long Short-Term Memory (LSTM) cells, and their effectiveness in improving the training stability of RNNs. |  |
|  | Explain the concept of Deep Recurrent Networks (DRNs) and their architecture for learning hierarchical representations of sequential data. Discuss the advantages of DRNs over shallow RNNs in capturing complex dependencies in long sequences. |  |
|  | Explain the fundamentals of Recurrent Neural Networks (RNNs) and their architecture for processing sequential data. Discuss the concept of recurrent connections and how RNNs maintain memory across time steps. |  |
|  | Analyze the architecture of Gated Recurrent Units (GRUs) and their mechanisms for handling the vanishing gradient problem in Recurrent Neural Networks (RNNs). Discuss the differences between GRUs and other recurrent units, such as LSTMs, and their relative effectiveness in different tasks. |  |
|  | a) Which model would you recommend for larger sequence data LSTM or RNN? Justify the choice. 2M  b) Give the values of Input, forget, output gates and values of updated cell states in LSTM. 3M |  |
|  | a) Compare RNN and Recursive neural networks(2M)  b)Explain encoder-decoder architectures of RNN with neat diagrams(3M) |  |
|  | Demonstrate the architectures of LSTM and GRU |  |
|  | Justify the reason behind Recurrent Neural Networks working better with text data |  |
|  | Assume that the memory cell in an Long Short-Term Memory architecture is supposed to sum its inputs over time, compute the values of input gate and forget gate. |  |
|  | Develop many to many RNN network for language translation with suitable figure and explanation. |  |
|  | Build LSTM network with all necessary logic gates. |  |
|  | Show with an example that, Vanishing gradient decreases the performances of RNN. |  |
|  | Explain various configurations of RNN with suitable diagram. |  |
|  | With near architecture, explain LSTM |  |
|  | Write on GRU and how RNN drawbacks are answered by GRUs |  |
|  | With neat diagram working of LSTM |  |
|  | Encoder decoder sequence to sequence architectures in context of RNNs |  |
|  | Sketch the graph of teacher forcing and networks with output recurrence and explain. |  |
|  | How the long-term dependencies challenge can be resolved in RNN.Explain |  |
|  | Compare LSTM and GRU |  |
|  | How RNNs can be represented as an Unfolded computational graphs. |  |
|  |  |  |
|  |  |  |
| UNIT-4 | | |
|  | What is an Autoencoder? | 2 |
|  | Define autoencoder with a neat diagram. | 2 |
|  | By using which method, Boltzmann machine reduces effect of additional stable states? | 2 |
|  | Mention generative model and its application. |  |
|  | Establish the relation between encoder, code and decoder |  |
|  | Describe about restricted boltzmans machine |  |
|  | What is sparse encoder and how does it work |  |
|  | Define RBMs |  |
|  | How RBMs are used for un-supervised learning. |  |
|  | How is Boltzmann machine different form deep Boltzmann machine |  |
|  | Difference between regular auto encoder and stochastic auto encoder. |  |
|  | ESSAY: |  |
|  | Analyze the architectural differences between Restricted Boltzmann Machines (RBMs) and traditional Boltzmann Machines. Discuss how RBMs are trained using the Contrastive Divergence algorithm and their applications in collaborative filtering tasks. |  |
|  | Evaluate the hierarchical structure of Deep Belief Networks (DBNs) composed of multiple layers of RBMs. Discuss the layer-wise pretraining approach used to train DBNs and their applications in image recognition and anomaly detection tasks. |  |
|  | Design a stochastic encoder for an autoencoder model and explain its significance in generating diverse representations of input data. Provide examples of tasks where stochastic encoders can enhance the performance of autoencoders. |  |
|  | Analyze the architectural differences between Restricted Boltzmann Machines (RBMs) and traditional Boltzmann Machines. Discuss how RBMs are trained using the Contrastive Divergence algorithm and their applications in collaborative filtering tasks. |  |
|  | Construct the steps involved in how a denoising autoencoder is trained to map a corrupted data point x1 to its original data point x. | 5 |
|  | Analyze the concept of undercomplete autoencoders and how they learn compressed representations of input data. Discuss the trade-offs of using undercomplete autoencoders and their applications in dimensionality reduction tasks. |  |
|  | Evaluate the role of regularization techniques in training autoencoders, focusing on methods like L1 and L2 regularization. Discuss how regularization helps prevent overfitting and improves the generalization performance of autoencoders. |  |
|  | Briefly write on deep belief networks? |  |
|  | Explain about sparse autoencoders and their generative model view. |  |
|  | Describe the different types of Boltzmann machines and their characteristics and architectures. |  |
|  | Justify the following statement: ‘Autoencoders can be used for feature generation’. |  |
|  | Illustrate the following with diagrammatic representation   1. Restricted Boltzmann Machine 2. Deep Belief Network   Deep Boltzmann Machine |  |
|  | Explain the architecture of Boltzmann Machine and prove that its computation cost is larger than RBM. |  |
|  | Demonstrate the uses of latent vector in Generative model with an example. |  |
|  | Apply Restricted Boltzmann Machine for movie recommendation system. Explain with architecture and example. |  |
|  | Conclude with necessary derivation that, noisy encoder is equivalent to L2 regularization. |  |
|  | Construct Generative Adversarial Networks with an explanation of generator and discriminator. Derive the cost function associated with this network. |  |
|  | Summarize the process used in for effective twitter data analysis. |  |
|  | What are DBMs and write on their chief properties and characteristics. |  |
|  | Write on auto encoders, sparse auto encoders and denoising auto encoders. |  |
|  | Problems faced in Deep belief networks. |  |
|  | How under complete auto encoder and decoder model learn a more generalization of PCA , explain |  |
|  | Problems in Deep belief networks |  |
|  | Working of denoising auto encoders |  |
|  | How do stochastic auto encoders different to standard auto encoders. Explain their working |  |
|  | Deep Boltzmann machine properites. |  |
|  |  |  |
|  |  |  |
| UNIT-5 | | |
|  | Given the sentence,” The man who does not read books has no advantage over the one who cannot read them”, frame the unigrams and bigrams. | 2 |
|  | List at least two possible applications of deep learning in NLP. | 2 |
|  | What is stemming in Natural Language Processing |  |
|  | Explain speech recognition |  |
|  | Mention three applications of computer vision |  |
|  | What significance does dataset augmentation hold in deep learning techniques |  |
|  | What are deep generative models |  |
|  | Compare exploration and exploitation |  |
|  | What is dataset augmentation |  |
|  | What is global contrast normalization |  |
|  |  |  |
|  | ESSAY: |  |
|  | What is contrast Normalization. Explain why local contrast normalization is preferred over global contrast normalization | 5 |
|  | a) Discuss the key advantages of GANs. 2M  b) Discuss various real time applications of GANs 3M |  |
|  | Explain how fast and distributed implementations are achieved in deep learning |  |
|  | Differentiate between GANs and Variational Autoencoders with architectures. |  |
|  | Explain the pre-processing techniques in Natural Language Processing |  |
|  | Differentiate Convolution Neural Networks from Generative Adversarial Networks |  |
|  | Mention the procedure of human emotion Recognition from the facial data. |  |
|  | Mention working principle of GAN in generative model. Represent the function of different agents associated in this model. |  |
|  | Deep learning techniques for the development of NLP |  |
|  | Define generative modelling. Write on working of GAN and its applications. |  |
|  | How deep learning techniques influence and drive computer vision research and applications. |  |
|  | Role of deep learning in speech recognition applications and algorithms. |  |
|  |  |  |