(2½ Hours)

[Total Marks: 60]

N. B.: (1) **All** questions are **compulsory**.
(2) Make **suitable assumptions** wherever necessary and **state the assumptions** made.
(3) Answers to the **same question** must be **written together**.
(4) Numbers to the **right** indicate **marks**.
(5) Draw **neat labelled diagrams** wherever **necessary**.
(6) Use of **Non-programmable** calculators is **allowed**.

I	**Choose the correct alternative and rewrite the entire sentence with the correct alternative.** **(30)**			
1.	Singular matrix are?			
	a.	non-invertible	**b.**	invertible
	c.	Both non-invertible and invertible	**d.**	
Answer : non invertible				
2.	Which of the following is not a type of matrix?			
	a.	Square Matrix	b.	Scalar Matrix
	c.	Trace Matrix	d.	Term Matrix
Answer : Term Matrix				
3.	What is true regarding Determinant of a Matrix?			
	a.	The concept of determinant is applicable to square matrices only.	**b.**	To find determinant, subtract diagonal elements together.
	c.	determinant is a vector value that can be computed from the elements of a Trace matrix	**d.**	A and C
Answer: The concept of determinant is applicable to square matrices only.				
4.	The concept of Eigen values and vectors is applicable to?			
	a.	Scalar matrix	b.	Identity matrix
	c.	Upper triangular matrix	d.	Square matrix
Answer: Square Matrix				
5.	Multiplication of a matrix with a scalar constant is called?			
	a.	Complex multiplication	**b.**	Linear multiplication
	c.	Scalar multiplication	**d.**	Constant multiplication
Answer: Scalar Multiplication				
6.	If A and B are parallel vectors then A × B =			
	a.	0	**b.**	1
	c.	-1	**d.**	2
Answer: 0				
7.	Which is the most direct application of neural networks?			
	a.	vector quantization	**b.**	pattern mapping
	c.	pattern classification	**d.**	control applications
Answer: pattern classification				
8.	How can false minima be reduced in case of error in recall in feedback neural networks?			
	a.	by providing additional units	**b.**	by using probabilistic update
	c.	can be either probabilistic update or using additional units	**d.**	by using vector update

Answer: by using probabilistic update				
9.	How many layers Deep learning algorithms are constructed?			
	a.	2	**b.**	3
	c.	4	**d.**	5
Answer: 4				
10.	What property should a feedback network have, to make it useful for storing information?			
	a.	accretive behaviour	**b.**	interpolative behaviour
	c.	both accretive and interpolative behaviour	**d.**	Accumulative behaviour
Answer: accretive behaviour				
11.	In Neural Network , the network capacity is defined as			
	a.	The traffic carry capacity of the network	**b.**	The total number of nodes in the network
	c.	The number of patterns that can be stored and recalled in a network	**d.**	The total number of layers in the network
Answer: The number of patterns that can be stored and recalled in a network				
12.	Which function is used to introduce non-linearity in Neural Network			
	a.	Transfer Function	**b.**	Error Function
	c.	Action Function	**d.**	Feed Forward Function
Answer: Transfer Function				
13.	The process performed in Convolution Layer called convolution operation is equivalent to which operation in Image Processing :			
	a.	Padding Operation	**b.**	Filter operation
	c.	Stride Operation	**d.**	Accumulation Operation
Answer: Filter operation				
14.	_____will apply element wise activation function to the output of convolution layer.			
	a.	Input Layer	**b.**	Convolution Layer
	c.	Activation Function Layer	**d.**	Pool Layer
Answer: Activation Function Layer				
15.	In _____ holds the raw input of image used to build ConvNets.			
	a.	Input Layer	**b.**	Convolution Layer
	c.	Activation Function Layer	**d.**	Pool Layer
Answer: Input Layer				
16.	.RNN Stands for _____.			
	a.	Recursive Neural Network	**b.**	Recurrent Neural Network
	c.	Recurring Neural Network	**d.**	Removable Neural Network
Answer: Recurrent Neural Network				
17.	The main and most important feature of RNN is _____.			
	a.	Visible State	**b.**	Hidden state
	c.	Present State	**d.**	Forward State
Answer: Hidden State				
18.	RNN remembers each and every information through_____.			
	a.	Work	**b.**	Time
	c.	Hour	**d.**	Memory
Answer: Time				
19.	_____ are algorithms or methods used to minimize error function or to maximize the efficiency of production			

	a.	Optimizers	**b.**	Activation function
	c.	Autoencoders	**d.**	Gradient Descent
Answer: Optimizers				
20.	Which algorithm uses different learning rates for each and every neuron for each and every hidden layer based on different iterations?			
	a.	AdaGrad	**b.**	RMS-Prop
	c.	AdaDelta	**d.**	Adam
Answer: AdaGrad				
21.	Which learning algorithm calculates adaptive learning rates for each parameter?			
	a.	Ada	**b.**	AdaDelta
	c.	Adam	**d.**	AdaDescent
Answer: Adam				
22.	Which algorithm was designed to handle the vanishing gradient problem?			
	a.	BRNN	**b.**	GRU
	c.	LSTM	**d.**	RMS
Answer: LSTM				
23.	Which networks are designed to handle vanishing gradient problem			
	a.	Graded Recurrent Units	**b.**	Long Short Term Memory
	c.	Bidirectional Recurrent Neural Network	**d.**	Long Term Memory
Answer: Graded Recurrent Units				
24.	Autoencoders consist of 3 parts :			
	a.	Input , Bottleneck, Decoder	**b.**	Encoder , Bottleneck, Output
	c.	Input, Bottleneck, Output	**d.**	Encoder, Bottleneck, Decoder
Answer: Encoder, Bottleneck, Decoder				
25.	___ _occurs when the gradients become very small and tend towards zero.			
	a.	Exploding Gradients	**b.**	Vanishing Gradients
	c.	Long Short Term Memory Networks	**d.**	Gated Recurrent Unit Networks
Answer : Vanishing Gradients				
26.	_____is responsible for reducing the spatial size of the Convolved Feature.			
	a.	Hidden layer	**b.**	Input Layer
	c.	Output Layer	**d.**	Pooling Layer
Answer : Pooling Layer				
27.	GAN stands for			
	a.	Generalized Adversarial Network	**b.**	Generative Adversarial Network
	c.	Generative Advanced Network	**d.**	Generative Artificial Network
Answer : Generative Adversarial Network				
28.	_____ is a neural network that identified real data from the fake data created by the Generator			
	a.	Discriminator	**b.**	Divider
	c.	Detector	**d.**	Determiner
Answer : Discriminator				
29.	_____ supports convolution neural networks instead of vanilla neural network at both Discriminator and Generator.			
	a.	Simple GAN	**b.**	Super Resolution GAN
	c.	Deep Convolution GANs	**d.**	Conditional GAN
Answer : Deep Convolution GANs				

30.	Which GANs have a min-max optimization formulation?			
	a.	Vanilla GAN	**b.**	Simple GAN
	c.	Super Resolution GAN	**d.**	Conditional GAN
Answer : Vanilla GAN				

Unit 1 Attempt any one of the following:

1. Write detailed note on Eigen Decomposition

Answer:

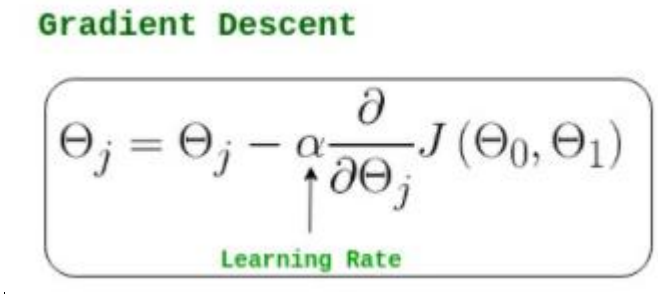
In linear algebra, Eigen decomposition is the factorization of a matrix into a canonical form, whereby the matrix is represented in terms of its eigenvalues and eigenvectors. Only diagonalizable matrices can be factorized in this way. When the matrix being factorized is a normal or real symmetric matrix, the decomposition is called "spectral decomposition", derived from the spectral theorem.

A (nonzero) vector **v** of dimension N is an eigenvector of a square $N \times N$ matrix **A** if it satisfies a linear equation of the form for some scalar λ. Then λ is called the eigenvalue corresponding to **v**. geometrically speaking, the eigenvectors of **A** are the vectors that **A** merely elongates or shrinks, and the amount that they elongate/shrink by is the eigenvalue.

2. What is Gradient base optimization?

Answer:

Gradient descent is an optimization algorithm that's used when training deep learning models. It's based on a convex function and updates its parameters iteratively to minimize a given function to its local minimum.



The notation used in the above Formula is given below, In the above formula,
α is the learning rate,
J is the cost function, and
Θ is the parameter to be updated.
As you can see, the gradient represents the partial derivative of J(cost function) with respect to Θj
Note that, as we reach closer to the global minima, the slope or the gradient of the curve becomes less and less steep, which results in a smaller value of derivative, which in turn reduces the step size or learning rate automatically.

It is the most basic but most used optimizer that directly uses the derivative of the loss function and learning rate to reduce the loss function and tries to reach the global minimum. Thus, the Gradient Descent Optimization algorithm has many applications including-

1. Linear Regression,
2. Classification Algorithms,
3. Backpropagation in Neural Networks, etc.

3. What is overflow and underflow?

Answer :

Deep Learning algorithms generally need a high volume of numerical computation. This normally states to algorithms that solve mathematical problems. That is solved by methods to keep informed guesses of the solution through an iterative process. Somewhat than logically deriving a formula in case a symbolic expression for the correct solution.

The general operations consist of optimization and solving systems of linear equations. As only assessing a mathematical function on a digital computer may be problematic when the function includes real numbers that cannot be signified exactly using a finite amount of memory. In this post, we will talk about Overflow and Underflow in detail.

Overflow and underflow are together errors ensuing from a shortage of space. They are distinct in data types similar to integers and floating points on the most basic level. The number kept in a computer occurs in a discrete number of digits different from the physical world.

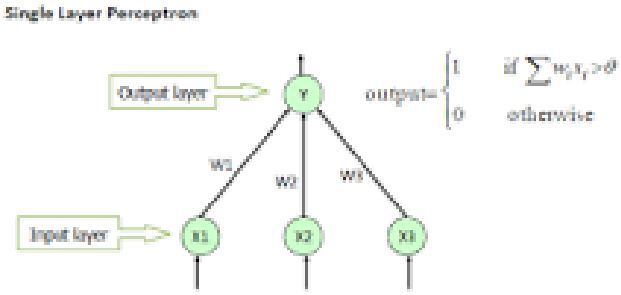
We cannot just append to our result when we make a calculation that results in an extra digit. Therefore, we acquire an overflow and underflow error. Overflow errors happen when functioning with integers and floating points.

Underflow errors are normally just related to floating points. The important trouble in performing continuous math on a digital computer is that we essential to embody considerably many real numbers with a finite number of minute designs. It means that for nearly all real numbers, we deserve some estimate error when we embody the number in the computer. This is an impartial rounding error in several cases.

Unit 2 Attempt any one of the following:

1. Write a detailed note on Single Layer Perceptron

Answer: Image result for Write a detailed note on Single Layer Perceptron. A single layer perceptron (SLP) is a feed-forward network based on a threshold transfer function. SLP is the simplest type of artificial neural networks and can only classify linearly separable cases with a binary target (1, 0).



2. What are Deep Feed Forward Networks explain with example?

Answer:

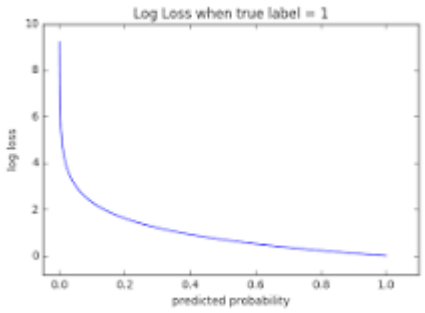
Deep feedforward networks, also often called feedforward neural networks, or multilayer perceptrons (MLPs), are the quintessential deep learning models. The goal of a feedforward network is to approximate some function f*. For example, for a classifier, y = f*(x) maps an input x to a category y. A feedforward network defines a mapping y = f(x;θ) and learns the value of the parameters θ that result in the best function approximation.(Reference)

These models are called feedforward because information flows through the function being evaluated from x, through the intermediate computations used to define f, and finally to the output y. There are no feedback connections in which outputs of the model are fed back into itself. When feedforward neural networks are extended to include feedback connections, they are called recurrent neural networks

3. Write a note on cross-entropy error

Answer :

Cross-entropy loss, or log loss, measures the performance of a classification model whose output is a probability value between 0 and 1. Cross-entropy loss increases as the predicted probability diverges from the actual label.



Unit 3 Attempt any one of the following:

1. What is backpropagation? What are various types of backpropagation networks?

Answer :

Backpropagation is the essence of neural network training. It is the method of fine-tuning the weights of a neural network based on the error rate obtained in the previous epoch (i.e., iteration). Proper tuning of the weights allows you to reduce error rates and make the model reliable by increasing its generalization.

Backpropagation in neural network is a short form for “backward propagation of errors.” It is a standard method of training artificial neural networks. This method helps calculate the gradient of a loss function with respect to all the weights in the network.

Types of Backpropagation Networks

Two Types of Backpropagation Networks are:
Static Back-propagation
Recurrent Backpropagation
Static back-propagation:
It is one kind of backpropagation network which produces a mapping of a static input for static output. It is useful to solve static classification issues like optical character recognition.
Recurrent Backpropagation:
Recurrent Back propagation in data mining is fed forward until a fixed value is achieved. After that, the error is computed and propagated backward.

2. Why a loss function is configured, explain with suitable example?
Answer:
The function used to evaluate a candidate solution (i.e. a set of weights) is referred to as the objective function. We may seek to maximize or minimize the objective function, meaning that we are searching for a candidate solution that has the highest or lowest score respectively. Typically, with neural networks, we seek to minimize the error. As such, the objective function is often referred to as a cost function or a loss function and the value calculated by the loss function is referred to as simply "loss."

3. Explain Convolution operations performed at convolution layer
Answer :
Convolutional layers are the major building blocks used in convolutional neural networks.
A convolution is the simple application of a filter to an input that results in an activation. Repeated application of the same filter to an input results in a map of activations called a feature map, indicating the locations and strength of a detected feature in an input, such as an image.
The innovation of convolutional neural networks is the ability to automatically learn a large number of filters in parallel specific to a training dataset under the constraints of a specific predictive modelling problem, such as image classification. The result is highly specific features that can be detected anywhere on input images.

Unit 4 Attempt any one of the following:

1. What are optimizers? Explain various types of optimizers
Answer :

Optimizers are algorithms or methods used to minimize an error function(*loss function*)or to maximize the efficiency of production. Optimizers are mathematical functions which are dependent on model's learnable parameters i.e Weights & Biases. Optimizers help to know how to change weights and learning rate of neural network to reduce the losses. Types of Optimizers

1. Gradient Descent
2. Stochastic Gradient Descent
3. Mini-Batch Gradient Descent
4. SGD with Momentum
5. AdaGrad(Adaptive Gradient Descent)
6. RMS-Prop (Root Mean Square Propagation)
7. AdaDelta
8. Adam(Adaptive Moment Estimation)

2 Explain the architecture of RNN

Answer :

There are two main architectures that are used in almost every application of recurrent neural networks: long-short term memory (LSTM) (Hochreiter and Schmidhuber, 1997) and gated recurrent unit (GRU) (Cho et al., 2014). Both of these use every time step to calculate an output and to update the internal state.

RNNs feed results back into the network (more on this point below). In CNNs, the size of the input and the resulting output are fixed.

3 Write a note on structure of LSTM

Answer:

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems. This is a behaviour required in complex problem domains like machine translation, speech recognition, and more. LSTMs are a complex area of deep learning.

Unit 5 Attempt any one of the following:

1. How does a Deep Generative Model works explain with example?

Answer :

Generative Adversarial Networks, or GANs for short, are an approach to generative modelling using deep learning methods, such as convolutional neural networks.

Generative modelling is an unsupervised learning task in machine learning that involves automatically discovering and learning the regularities or patterns in input data in such a way that the model can be used to generate or output new examples that plausibly could have been drawn from the original dataset.

GANs are a clever way of training a generative model by framing the problem as a supervised learning problem with two sub-models: the generator model that we train to generate new examples, and the discriminator model that tries to classify examples as either real (from the domain) or fake (generated). The two models are trained together in a zero-sum game, adversarial, until the discriminator model is fooled about half the time, meaning the generator model is generating plausible examples.

GANs are an exciting and rapidly changing field, delivering on the promise of generative models in their ability to generate realistic examples across a range of problem domains, most notably in image-to-image translation tasks such as translating photos of summer to winter or day to night, and in generating photorealistic photos of objects, scenes, and people that even humans cannot tell are fake

2. What is a Discriminator in Generative Models?

GANs are a clever way of training a generative model by framing the problem as a supervised learning problem with two sub-models: the generator model that we train to generate new examples, and the discriminator model that tries to classify examples as either real (from the domain) or fake (generated)

3.Write a note on Deep Convolution GANs.

DCGAN, or Deep Convolutional GAN, is a generative adversarial network architecture. It uses a couple of guidelines, in particular: Replacing any pooling layers with strided convolutions (discriminator) and fractional-strided convolutions (generator). Using batch norm in both the generator and the discriminator.
