USN

Internal Assessment Test 2 Answer Scheme– November 2024

Sub:	DEEP LEARNING					Sub Code:	21CS743	Branch:	AInDS	
Date:	20/11/2024	Duration:	90 minutes	Max Marks:	50	Sem	VII			OBE

		Answer any FIVE Questions	MARKS	CO	RBT
1	a	**Analyze and write short notes on Dataset Augmentation and its applications.** The best way to make a machine learning model generalize better is to train it on more data. Of course, in practice, the amount of data we have is limited. One way to get around this problem is to create fake data and add it to the training set **(3 Marks)** Applications with Explanation : object recognition , speech recognition etc **(3 Marks)**	[6]	CO 2	L3
	b	**Explain Back Propagation algorithm** The back-propagation algorithm often simply called backprop, allows the information from the cost to then flow backwards through the network, in order to compute the gradient**. (2 Marks)** back-propagation refers only to the method for computing the gradient and how to compute the gradient $\nabla_x f(x, y)$ for an arbitrary function f **(2 marks)**	[4]	CO 2	l2
2	a	**Discuss in detail about chain rule of calculus.** The chain rule of calculus (not to be confused with the chain rule of probability) is used to compute the derivatives of functions formed by composing other functions whose derivatives are known. Back-propagation is an algorithm that computes the chain rule, with a specific order of operations that is highly efficient**. (2 Marks)** **Examples : (3 Marks)**	[5]	C0 2	L2
	b	**Explain Multi-Task Learning and semi supervised learning** MTL pertains to instructing a neural network to undertake several tasks, achieved by distributing certain network layers and parameters across these tasks. **(2 Marks)** The model can generally be divided into two kinds of parts and associated parameters: 1. Task-specific parameters 2. 2. Generic parameters **(1 Mark)** Semi-supervised learning is a branch of machine learning that combines supervised and unsupervised learning by using both labeled and unlabeled data to train artificial intelligence (AI) models for classification and regression tasks. Eg : Image and Speech Analysis **(2 Marks)**	[5	C0 2	
3	a	**Illustrate Computational graphs with examples** The back-propagation algorithm more precisely, it is helpful to have a more precise computational graph language. To formalize our graphs, we also need to introduce the idea of an operation. An operation is a simple function of one or more variables. Our graph language is accompanied by a set of allowable operations. If a variable y is computed by applying an operation to a variable x, then we draw a directed edge from x to y **(3 Marks)** Example **(2 Marks)**	[5]	C0 2	L3

	b	**Discuss about Adversarial Training with Examples.** Adversarial Training is a machine learning technique that is primarily used for improving the robustness of models. It's a process where models are trained with malicious inputs alongside the genuine data. **(2 Marks)** Adversarial **examples** also provide a means of accomplishing semi-supervised learning. At a point x that is not associated with a label in the dataset, the model itself assigns some label yˆ. The model's label yˆ may not be the true label, but if the model is high quality, then yˆ has a high probability of providing the true label. (**3 Marks)**	[5]		
4		**Explain about Adagrad and RMSProp Algorithm in Detail.** **Adagard :** The AdaGrad algorithm, shown in algorithm 8.4, individually adapts the learning rates of all model parameters by scaling them inversely proportional to the square root of the sum of all of their historical squared values . The parameters with the largest partial derivative of the loss have a correspondingly rapid decrease in their learning rate, while parameters with small partial derivatives have a relatively small decrease in their learning rate. The net effect is greater progress in the more gently sloped directions of parameter space. **(2 Marks)** Algorithm **: (3 Marks)** **RMSProp :** RMSProp uses an exponentially decaying average to discard history from the extreme past so that it can converge rapidly after finding a convex bowl, as if it were an instance of the AdaGrad algorithm initialized within that bowl. **(2 Marks)** Algorithm **: (3 Marks)**	[10]	C0 2	L2
5	a	**Explain Stochastic Gradient Descent (SGD)** Stochastic Gradient Descent (SGD) is a powerful optimization algorithm used in machine learning and artificial intelligence to train models efficiently. It is a variant of the gradient descent algorithm that processes training data in small batches or individual data points instead of the entire dataset at once**.**	[5]	C0 2	L2
	b	**Explain regularization for deep learning.** Many strategies used in machine learning are explicitly designed to reduce the test error, possibly at the expense of increased training error. These strategies are known collectively as regularization**(2 Marks)** An effective regularizer is one that makes a profitable trade, reducing variance significantly while not overly increasing the bias**(1 Mark)** we focused on three situations, where the model family being trained either (1) excluded the true data generating process—corresponding to underfitting and inducing bias, or (2) matched the true data generating process, or (3) included the generating process but also many other possible generating processes—the overfitting regime where variance rather than bias dominates the estimation error. The goal of regularization is to take a model from the third regime into the second regime. **(2 Marks)**	[5]	C0 2	
6	a	**Define Bagging and Discuss in detail about Dropout.** Bagging : is a technique for reducing generalization error by combining several models. he idea is to train several different models separately, then have all of the models vote on the output for test examples. This is an example of a general strategy in machine learning called model averaging. Techniques employing this strategy are known as ensemble methods. **(2 Marks)** Dropout : Dropout is a technique where randomly selected neurons are ignored during training. They are "dropped-out" randomly. This means that their contribution to the activation of downstream neurons is temporally removed on the forward pass and any weight updates are not applied to the neuron on the backward pass.(removing non-output units from an underlying base network). To train with dropout, we use a minibatch-based learning algorithm that makes small steps, such as stochastic gradient descent. **(3 Marks)**	[5]	C0 2	L2

	b	**Explain about Surrogate Loss Function and Early Stopping.** The idea is to use the surrogate loss function during training to update the model parameters, and then evaluate the performance of the model using the original loss function. Surrogate loss functions are alternative loss functions that are used in machine learning when the original loss function is difficult or impossible to optimize directly. One of the main benefits of using surrogate loss functions is simplicity. The original objective function may be very complex and difficult to optimize, while a surrogate function can be designed to be simpler and more computationally efficient. This simplicity can make it easier to optimize the function and reduce the computational cost of training the model. **(3 Marks)** **Early stopping :** Early stopping is a powerful technique for training deep learning models. It strikes a balance between underfitting and overfitting, ensuring the model generalizes well. By monitoring the validation loss and halting training at the right moment, early stopping prevents overfitting and saves computational resources. **(2 Marks)**	[5]	C02	

CI CCI HOD