|  |  |  |  |
| --- | --- | --- | --- |
| **Sl. No** | **Test Questions** | **CO** | **Bloom’s**  **Level** |
| **1** | In a binary classification problem use a Bayesian approach to update your belief about the parameter 0, which represents the probability of classifying an instance as positive (class 1). The prior distribution for is a Beta distribution with parameters a=3 and B-5.  You observe the following data:  10 positive instances (class 1)  6 negative instances (class 0)  Calculate the following:  a) Determine the likelihood of observing the given data given the parameter 0. (3 marks)  b) Find the posterior distribution of after observing the data. (3 marks)  c) Calculate the expected value of the posterior distribution of 0. (4 marks) | **CO1** | **K3** |
| **2** | Three persons A, B and C have applied for a job in a private company. The chance of their selections is in the ratio 1 : 2 : 4. The probabilities that A, B and C can introduce changes to improve the profits of the company are 0.8, 0.5 and 0.3, respectively. If the change does not take place, find the probability that it is due to the appointment of C. | **CO1** | **K2** |
| **3** | Consider a deep learning model with the following training and validation loss curves during training:  • Epoch 1: Training Loss = 2.0, Validation Loss = 1.8  • Epoch 2: Training Loss = 1.8, Validation Loss = 1.7  • Epoch 3: Training Loss = 1.7, Validation Loss = 1.9  • Epoch 4: Training Loss = 1.5, Validation Loss = 2.5  • Epoch 5: Training Loss = 1.3, Validation Loss = 3.0  • Epoch 6: Training Loss = 1.2, Validation Loss = 3.5  • Epoch 7: Training Loss = 1.1, Validation Loss = 4.0  • Epoch 8: Training Loss = 1.0, Validation Loss = 4.5  a) Identify the point at which the model starts to overfit.  b) Describe a strategy to address overfitting in this scenario.  c) If the training loss continued to decrease while the validation loss kept increasing, what could be the consequence for the model's performance on  unseen data? | **CO1** | **K3** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl. No** | **Test Questions** | **CO** | **Bloom’s**  **Level** |
| **1** | Suppose you are training a convolutional neural network (CNN) for image classification on a dataset containing 10.000 images. You have decided to use a validation set to tune your hyperparameters. You split your dataset into 80% training data and 20% validation data.  Your CNN architecture consists of the following hyperparameters  1. Learning rate (a) 0.001 to 0.01 in steps of 0.001  2. Dropout rate (p) 0.1 to 0.5 in steps of 0.1  3. Number of filters (F) 16. 32 or 64  4. Size of the kernel (K): 3x3 or 5x5  5. Number of hidden units in fully connected layers (H) 256 or 512  You decide to perform a grid search over these hyperparameters to find the combination that yields the best validation accuracy  a) How many unique hyperparameter combinations are you testing in total? (4 Marks)  b) If each training epoch takes 10 minutes and you want to limit your grid search to a maximum of 100 epochs per hyperparameter combination, how much time will it take to complete the entire grid search? (4 Marks)  c) Suggest a strategy to reduce the time taken for hyperparameter tuning while still ensuring a good selection of hyperparameters (2 Marks) | **CO2** | **K3** |
| **2** | Consider a Convolutional Neural Network (CNN) architecture for image classification with the following specifications  First convolutional layer 32 filters of size 3x3, with stride 1 and no padding.  Input image size 128x128 pixels. - Max pooling layer after the first convolution: 2x2 pooling window with stride 2 Second convolutional layer 64 filters of size 5x5, with stride 1 and padding to maintain the spatial dimensions  - Max pooling layer after the second convolution: 2x2 pooling window with stride 2 - Fully connected layer with 128 neurons  - Output layer with softmax activation for classification into 10 classes. Given an input image, how many parameters (weights and biases) are there in the convolutional layers and fully connected layers of this CNN? | **CO2** | **K3** |
| **3** | A insurance company has insured 4000 doctors, 8000 teachers and 12000 businessmen. The chances of a doctor, teacher and businessman dying before the age of 58 is 0.01, 0.03 and 0.05, respectively. If one of the insured people dies before 58, find the probability that he is a doctor | **CO1** | **K2** |
| **Sl. No** | Test Questions | **CO** | **Bloom’s**  **Level** |
| **1** | Suppose you have a dataset consisting of sequences of length 10, where each element in the sequence is a one-hot encoded vector representing a character from an alphabet of size 26 (a-z). You want to train an RNN for sequence prediction on this dataset  a) If you use a basic RNN architecture with 64 hidden units, how many  parameters (weights and biases) does the RNN model have? b) Assuming each one-hot encoded vector takes 26 floats to represent, how  many total input features are there for each time step in the sequence?  c) If you have 1000 sequences in your training dataset, each of length 10, how many total training samples (input-output pairs) do you have?  d) Assuming a softmax output layer, if you want to predict the next character in the sequence, how many output neurons are needed in the final layer of the RNN model? | **CO3** | **K3** |
| **2** | Let's consider a dataset of time series measurements with 100 features, sampled over 500 time steps. You want to apply Slow Feature Analysis to extract the slowest features from this dataset. Assuming you choose to extract the top 5 slowest features.  a) If you use a polynomial expansion of degree 2 to create the input space for SFA, how many input features will you have after the expansion?  b) Assuming you use a window size of 10 for temporal differencing in SFA, what will be the length of the resulting time series after differencing?  c) If you apply a whitening step to the data before performing SFA and choose to keep 90% of the variance, how many principal components will be retained?  d) After performing SFA, if you want to visualize the learned features, how many dimensions will the feature space have if you choose to represent it using t-SNE with a perplexity of 30? | **CO3** | **K3** |
| **3** | When a company implements advertisements on two of their prime products, the revenue (Y) increases. Let the products are A and B. When an analysis is done to increase the revenue by combining these product sales; it provides the following relation. Compute whether this task is linearly separable or not. | **CO2** | **K2** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl. No** | **Test Questions** | **CO** | **Bloom’s**  **Level** |
| **1** | At a comprehensive level, can you evaluate the significance of hyperparameter selection in the training of neural networks? Further, analyze the methodologies of grid search and random search as prevalent techniques utilized to ascertain the optimal hyperparameters, emphasizing their respective processes, advantages, and limitations within the context of neural network optimization. | **CO4** | **K2** |
| **2** | In an industrial process, an AI robot is assigned to segregate the front and back tires automatically. Let the number of tires are 250 (125 front and 125 back tires). • Front identified as Front: 119 • Front identifies as Back: 06 • Back identified as Back: 120 • Back identified as Front: 05 Compute: Accuracy, Precision, Sensitivity, Specificity and F1-Score. | **CO4** | **K3** |
| **3** | During an experimental investigation, a mathematical model is achieved as a function of; 𝑓 Ө = 𝑓 𝑥, 𝑦, 𝑧 = 𝑥 2 + 𝑦 2 + 𝑧 2 − 4. What will be the new value of the parameters, if the gradient of this function at Ө = {2, −2, 1}. Note that Ө = {𝑥, 𝑦, 𝑧} is a collection of all the parameters of this function. (a) [4, -4, 2]. (b) [-2, 4, -4], (c) [4, -4, -2], (d) [-1, -4, 4]. | **CO4** | **K2** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl. No** | **Test Questions** | **CO** | **Bloom’s**  **Level** |
| **1** | Consider a deep neural network with the following architecture for a binary classification task:  Input layer with 100 features  Two hidden layers with 200 and 100 neurons respectively, both using ReLU activation  Output layer with a single neuron using Sigmoid activation  The network is trained using stochastic gradient descent (SGD) with a batch size of 32 and a learning rate of 0.01. The loss function used is binary cross-entropy.  a) Calculate the total number of parameters (weights and biases) in this network.  b) If the training dataset contains 10,000 samples, how many times will the model parameters be updated in one epoch?  c) If the training process converges after 100 epochs, how many parameter updates would have been made in total? | **CO2** | **K3** |
| **2** | During a classification task, an unbalanced data with a ratio of 5:4 was considered to train the deep-learning scheme and achieved the following results; True Positive (TP) of 91 for a testing data size of 100 and a True Negative (TN) of 69. Calculate the Accuracy and F1-Score. | **CO3** | **K3** |
| **3** | A machine learning model is being trained on a dataset with 1000 samples. The dataset is divided into training, validation, and test sets with a ratio of 70:15:15. The model has two hyperparameters: learning rate and regularization strength.  The learning rate is chosen from the set {0.001, 0.01, 0.1}.  The regularization strength is chosen from the set {0.01, 0.1, 1.0}.  If we use grid search to find the best combination of hyperparameters using the validation set, how many different combinations of hyperparameters do we need to try? | **CO3** | **K2** |