EN 3150 Model Questions.

Question 1.

(a) Suppose you have a huge data set and you have time constrains. Here, which variation of the gradient descent algorithm would you select and what is the primary reason behind your selection? [02 marks]

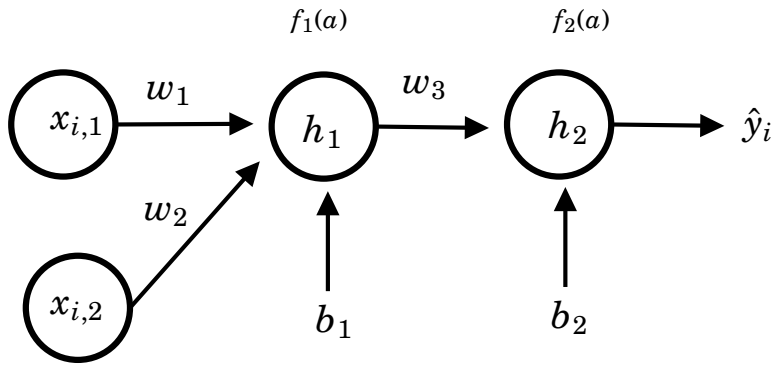

Figure Q1.1: Simple Network.

(b) Backpropagation and stochastic gradient decent are the two main operations in training a neural network. Consider the simple network shown in Figure Q1.1. Stating form $x_{1,1} = 5$ and $x_{1,2} = 4$ run one iterations of stochastic gradient descent with a learning rate of 0.1. Initialize weights $w_1 = w_2 = w_3 = 1$ and take true value $y_1 = 1$ and $b_1 = b_2 = 0.1$. Here, ReLU ($f_1(a)$) and Sigmoid ($f_2(a)$) activation functions are used. For the loss function, binary cross entropy loss function is used. It is given by $l(\boldsymbol{w}) = \frac{-1}{N}\sum_{i=1}^{N}(y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i))$. What are the values of the w_1, w_2 and w_3 after one iteration. [15 marks]

(c) Suppose that you trained a support vector machine (SVM) using a data set and record the test accuracy. Next, you randomly removed 70% of training data and retained the SVM model and test with the same testing data set. You have seen that the testing accuracy does not change. What could be the possible reason for this. [04 marks]

(d) Suppose that probability of having a disease is 2%. A test is used to detect this disease and it is known to be 90% accurate in detecting it when you have it and 95% accurate in confirming that you do not have it when you do not have the disease. If someone takes this test and it comes back positive, what is the probability that the person actually has the disease? [04 marks]

. .

Question 2.

(a) A data set of five data samples are given in Table Q2.1 . Suppose that the regression model $f(\boldsymbol{x}) = y = w_0 + w_1 \frac{x_2}{x_1}$ fits to the given data set. Use all the data samples to find parameters of the regression model (w_0 and w_1). [10 marks]

Table Q2.1: Data set for Q2.(a).

Sample index (i)	$x_{1,i}$	$x_{2,i}$	y_i
1	5	5	7
2	5	10	12
3	4	12	17
4	$\frac{25}{4}$	25	22
5	$\frac{1}{3}$	$\frac{5}{3}$	27

(b) A logistic regression model has been trained on a dataset of emails. Suppose that the independent features are length of the email (x_1), number of times the word "free" appears in the email (x_2), and number of times the word "hurry" appears in the email (x_3). The dependent feature (y) represents whether an email is spam ($y = 1$) or not spam ($y = 0$). After training, the logistic regression coefficients are as follows: $w_0 = 0.05$, $w_1 = 0.001$, $w_2 = 0.25$, and $w_3 = 0.1$. The logistic regression model is given by $p(y = 1|\boldsymbol{x}, \boldsymbol{w}) = \text{sigm}(w_0 + w_1x_1 + w_2x_2 + w_3x_3)$ and sigm$(\cdot)$ is the sigmoid function. What is the probability of an email is a spam email, when $x_1 = 100$, $x_2 = 2$, and $x_3 = 1$? [05 marks]

(c) State two advantages of generative classifiers over discriminative classifiers? [02 marks]

(d) Briefly explain the k-means algorithm cluster assignments and cluster centers update process. [05 marks]

(e) Suppose that the data is not linearly separable, meaning that a linear boundary cannot effectively separate the classes. In this scenario, name three approaches that can be employed to classify this data. [03 marks]

. .

Question 3.

(a) Suppose you are given a task to design a deep neural network for the following scenarios. Your goal is to decide the dimension of the final layer in each case.

i. Classification of handwritten digit which has 10 classes.

ii. Classification of flower which has 3 classes.

iii. Predict the selling price of a house.

[03 marks]

(b) Consider a convolutional neural network (CNN) implementation for recognizing facial expressions in images. The images are 24×24 RGB (3 color channels). Suppose that there are five different facial expressions to recognize. The architecture of the CNN is given in Figure Q3.1.

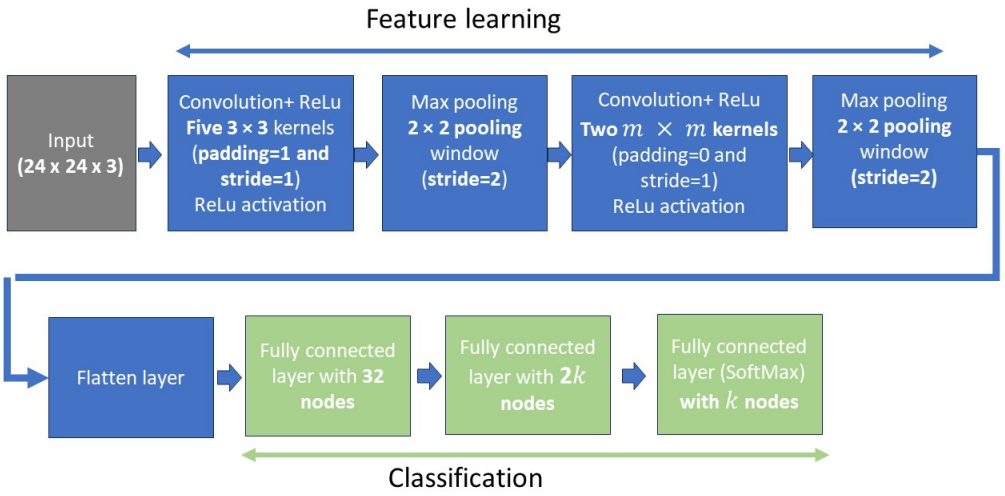

Figure Q3.1: CNN for recognizing facial expressions.

i. Determine m and k. [06 marks]

ii. Compute the total number of learnable parameters (weights and biases) of this CNN. [06 marks]

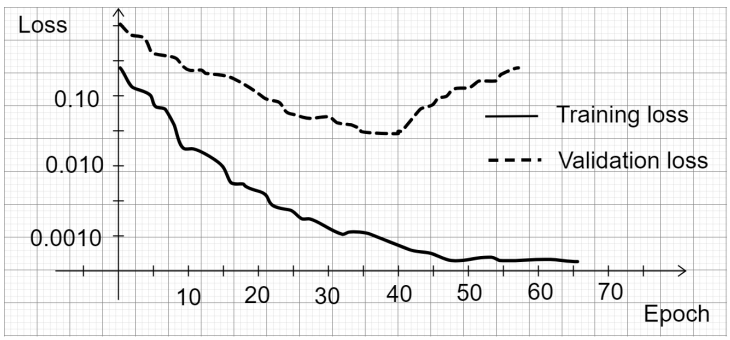

Figure Q3.2: Loss function of CNN training.

(c) Given the loss function plotted against the number of epochs in Figure Q3.2, at what point do you decide to stop the training of the CNN, and what is the reason this decision? [05 marks]

(d) Assume that there is a similar dataset for 10 facial expressions. Describe how this trained network can be used to classify 10 facial expressions. [05 marks]

. .

Question 4.

(a) Training accuracy of a given machine learning model before feature standardization is 70% and after feature standardization, the accuracy is dropped to 40%. What could be the possible reason? [05 marks]

(b) Suppose you have a machine learning model which is initially trained on handwritten digit images. You need to use this model to detect noisy handwritten digit images. Propose a method to enhance the image quality and reduce noise in these images and briefly explain the process. [05 marks]

(c) Decision surface of a dataset is shown in Figure Q4.1. Draw a decision tree based on the given decision surface. [08 marks]

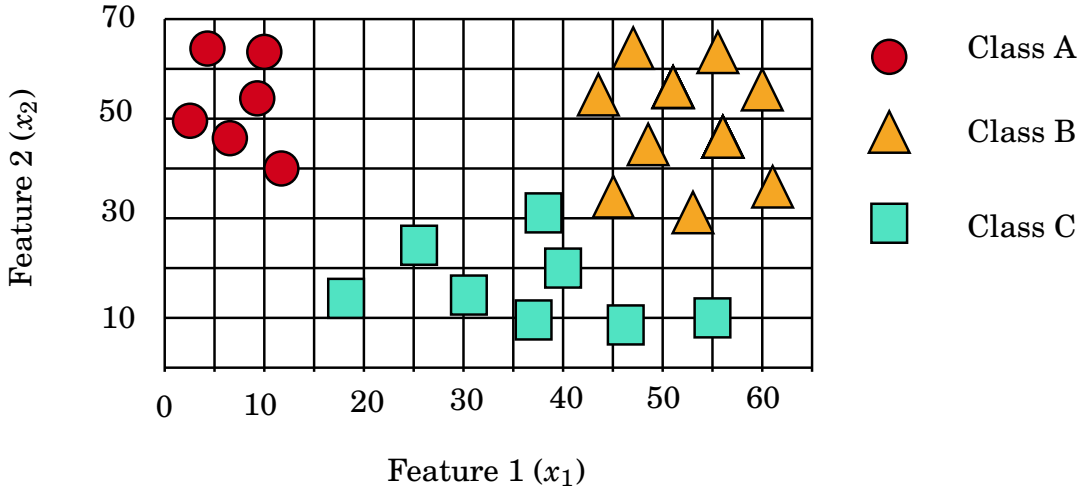

Figure Q4.1: Decision surface of a dataset.

(d) It is given that $k_1(\boldsymbol{x},\ \boldsymbol{z}) = \left(\boldsymbol{x}^T\boldsymbol{z}\right)$ is a valid kernel function. Is $k(\boldsymbol{x},\ \boldsymbol{z}) = \exp\left(\boldsymbol{x}^T\boldsymbol{z}\right)$ a valid kernel function? Justify your answer. [05 marks]

(e) State one application of kernel PCA. [02 marks]

. .