Lebanese University
Faculty of information
Section 1

Final Exam: Machine Learning

Time: 90 minutes
Dr. Abbas RAMMAL

Problem 1: Neural Network Forward & Backpropagation

A neural network is used to predict the probability of disease progression based on three standardized clinical features. The input feature vector is:

where:

$$x = \begin{bmatrix} 0.8 \\ -0.4 \\ 1.2 \end{bmatrix}$$

x_1: Age (standardized)

x_2: Body Mass Index (standardized)

x_3: Biomarker level (standardized)

The hidden layer consists of three neurons with ReLU activation.

- Weights: $W^{(1)} = \begin{pmatrix} 0.2 & -0.5 & 0.3 \\ -0.4 & 0.1 & 0.6 \\ 0.7 & -0.2 & 0.1 \end{pmatrix}$
- Bias: $b^{(1)} = \begin{bmatrix} 0.1 \\ -0.2 \\ 0.0 \end{bmatrix}$
- Activation function: ReLU(z) = max(0, z)

The output layer consists of **one neuron** with **sigmoid activation.**

- Weights: $W^{(2)} = [0.5 \ -0.7 \ \ 0.9]$
- Bias: $b^{(2)} = -0.3$
- Activation function: $\sigma(z) = \frac{1}{1+e^{-z}}$

Assume that the true target value for this observation is: $y = 1$. The network is trained using the Mean Squared Error (MSE) loss function.

- Learning Rate : $\eta = 0.01$

Questions

1. Construct and clearly label the architecture of the neural network, indicating:
   - Number of input features
   - Number of hidden neurons

- Activation functions
- Output neuron

2. Compute the hidden layer activation vector.
3. Compute the final output $\hat{y}$ of the network.
4. Using the MSE loss function, compute:
   - the loss value
   - the error term at the output neuron
5. Propagate the error from the output layer back to the hidden layer
6. Using gradient descent:
   - Update all weights in the output layer
   - Update all weights in the hidden layer

Problem 2. Regression model with Regularization

You're training a ridge regression model on a batch of two samples with the loss function:

$$L(w,b) = \frac{1}{2}\sum_{i=1}^{n}(\hat{y_i} - y_i)^2 + \lambda\|w\|_2^2$$

Given:

- Samples: (x_1 = [2, 1], y_1 = 4), (x_2 = [1, 3], y_2 = 7)
- Initial weights: w = [0, 0], b = 0
- Learning rate $\eta = 0.1$, regularization $\lambda = 1$

a) Derive the gradient of the loss with respect to w and b for the batch.
b) Perform one batch gradient descent update. Show all steps.
c) Explain the trade-off between underfitting and overfitting when tuning λ.

Problem 3. Naïve Bayes Classifier

Let's consider a dataset with a categorical target variable and two categorical predictor variables. Let's say we want to predict whether to play based on the outlook (Sunny, Overcast, Rainy) and temperature (Hot, Mild, Cool).

Instance	Outlook	Temperature	Play
1	Sunny	Hot	No
2	Sunny	Hot	No
3	Overcast	Hot	Yes

4	Rainy	Mild	Yes
5	Rainy	Cool	Yes
6	Rainy	Cool	No
7	Overcast	Cool	Yes
8	Sunny	Mild	Yes

1. Using Naive Bayesian classification method, predict a class label *(yes or no) for the* following unknown's samples:
   - S1: (Sunny, Mild) Y
   - S2: (Sunny, Cool) N
   - S3: (Sunny, Hot) N
   - S4: (Rainy, Mild) Y
   - S5: (Overcast, Hot) Y
   - S6: (Overcast, Mild) Y

   Show all computation steps.

2. The actual class labels for the unknown samples are as follows:
   - S1: No
   - S2: Yes
   - S3: No
   - S4: Yes
   - S5: Yes
   - S6: No

a) Construct the confusion matrix to compare the predicted class labels with the actual class labels for all samples.

b) Calculate various evaluation metrics such as accuracy, precision, recall, and F1-score to assess the performance of the classifier.