Deep Learning Assignment

Ganesh Birua
Roll - 21108

Q.1

Cross-entropy (CE) and Mean Squared Error (MSE) are both loss functions used for supervised learning algorithms. In the case of logistic regression, cross-entropy guarantees the best answer. Cross-entropy is a measure of the difference between two probability distributions for a given random variable or set of events. The cross-entropy between two probability distributions, such as Q from P, can be stated formally as:

$$H(P,Q)$$

Where $H()$ is the cross-entropy function, P may be the target distribution, and Q is the approximation of the target distribution. Cross-entropy can be calculated using the probabilities of the events from P and Q, as follows:

$$H(P,Q) = -\sum_{x \in X} P(x) \cdot \log(Q(x))$$

By quantifying this disparity, it guides the training process, enabling parameter updates via gradient descent algorithms. Specifically, in classification models, it fosters accurate probability estimation through techniques like softmax activation, enhancing the model's ability to discern between classes. Moreover, cross-entropy serves a regularization role, helping prevent overfitting by penalizing overly complex models. Its effective use ensures model convergence towards optimal predictions while minimizing divergence from ground truth distributions, thereby playing a vital role in achieving robust and accurate machine learning models.

Q.2

For binary classification tasks with deep neural networks employing linear activation functions, the Cross-Entropy (CE) loss function ensures convex optimization. CE loss, widely used in binary classification, is formulated to penalize the divergence between predicted probabilities and true labels. Analytically, CE loss exhibits convexity, as verified by the second derivative. Linear activation functions result in linear decision boundaries, making the optimization problem convex. With linear decision boundaries and convex loss, the optimization problem becomes convex, guaranteeing convergence to the global minimum. Hence, for binary classification tasks with linear activation functions, CE loss offers convex optimization, ensuring efficient and reliable model training.

Q.3

The model consists of a feedforward neural network with three dense (fully connected) layers. These are the details:

Number of Hidden Layers: The model has two hidden layers.
The first hidden layer has 256 neurons.
The second hidden layer has 128 neurons.

Activation Functions:
Both hidden layers use the Rectified Linear Unit (ReLU) activation function.
The output layer uses the softmax activation function. It outputs probabilities for each class, and the class with the highest probability is predicted.

Q.4

Below are the results obtained while running the architectures.
textbfTraining and evaluating VGG-16...

```
Epoch 1/5, Loss: 2.288294100595268
Epoch 2/5, Loss: 1.058486030284536
Epoch 3/5, Loss: 0.49857990267177077
Epoch 4/5, Loss: 0.35887506611982706
Epoch 5/5, Loss: 0.2855098379795352
Accuracy: 0.9094960049170252
```

Training and evaluating ResNet-18...

```
Epoch 1/5, Loss: 2.0546019673347473
Epoch 2/5, Loss: 0.7374378602679182
Epoch 3/5, Loss: 0.5432222432064262
Epoch 4/5, Loss: 0.3694945172981103
Epoch 5/5, Loss: 0.29688645213946235
Accuracy: 0.8845267363245236
```

Training and evaluating ResNet-50...

```
Epoch 1/5, Loss: 2.3068630649653046
Epoch 2/5, Loss: 1.5649826250425198
Epoch 3/5, Loss: 1.01850483984482
Epoch 4/5, Loss: 0.8195795713816786
Epoch 5/5, Loss: 0.7497649172041889
Accuracy: 0.7956745543945912
```

Training and evaluating ResNet-101...

```
Epoch 1/5, Loss: 2.6094645938806833
Epoch 2/5, Loss: 2.429790206902534
Epoch 3/5, Loss: 2.3804311278805086
Epoch 4/5, Loss: 2.3331697888490632
Epoch 5/5, Loss: 2.094680352078082
Accuracy: 0.4108020897357099
```

Comments on model performance:

VGG-16 and ResNet-18 models achieve the highest accuracies, with VGG-16 performing slightly better. This is likely because both VGG-16 and ResNet-18 are relatively shallow architectures compared to ResNet-50 and ResNet-101, making them more suitable for the SVHN dataset which might not require very deep networks for effective feature extraction. ResNet-50 and ResNet-101, being deeper architectures, may suffer from overfitting on the SVHN dataset due to their increased complexity and capacity. Additionally, the SVHN dataset may not have sufficiently complex features to benefit from the deeper architectures.