Indian Institute of Technology Tirupati
Deep Learning (CS5223): End Semester

Date: 9^{th} **December, 2021 (2 - 5 PM)**
Maximum Marks: 60
Instructions: Answer all the questions.

---

1. An Autoencoder has the following two modules: $\tilde{X} = Decoder(Z, W_d)$ and $Z = Encoder(X, W_e)$ where W_e, W_d are the parameters of encoder and decoder.
   (a) Given a training set $\{X_i\}$ comprising of N data points in D dimensions, provide a possible loss function for training the autoencoder. [2 Marks]
   (b) How does the loss function change in a denoising autoencoder to prevent the system from converging to a trivial and useless solution? [2 Marks]

2. (a) Draw a basic Geneative Adversarial Network (GAN) using the elements corresponding to its two players, source of real images, and a source of randomness. [2 Marks]
   (b) Write the expression for the objective functions of both the players. Clearly define all the terms involved. [2 Marks]

3. Give two benefits of using convolutional layers instead of fully connected ones for computer vision tasks. [2 Marks]

4. You are solving the binary classification task of classifying images as cat vs. non-cat. You design a CNN with a single output neuron. Let the output of this neuron be z. The final output of your network, $\hat{y}$ is given as
   $$\hat{y} = \sigma(ReLU(z))$$
   You classify all inputs with a final value $\hat{y} \geq 0.5$ as cat images. What problem are you going to encounter? [2 Marks]

5. Why is softmax operation a special kind of nonlinearity? How is it different from other activation functions (apart from the function)? [2 Marks]

6. Archaeological department approached you for developing a deep learning based classifier for categorizing historical objects into three categories: Antiquity, Middle-age, and Modern. Over years the department have collected about 5000 human labelled RGB images. These are sample images from their dataset (don't bother if you are looking at their gray-scale versions).

Figure 1: Class 0 - Antiquity

Figure 2: Class 1 - Middle-age

Figure 3: Class 2 - Modern

   (a) Before training your model, you want to decide the image resolution to be used. Why is the choice of image resolution important? [2 Marks]
   (b) If you have very small time to decide on the resolution, what would you do? [2 Marks]
   (c) After you figured out a resolution, how would you partition your dataset (you may formulate in percentages)? [2 Marks]
   (d) After visually inspecting your partitions, you come to know that the training set only contains pictures taken during the day, whereas the validation set only has images taken at night. Is there

an issue here? If so, how you would correct it? [2 Marks]
(e) If you feel that the training data is not enough and you can not get more images until next year, what would you do with the available dataset to improve the performance of your DNN classifier? (mention at least 3 tricks you can apply) [2 Marks]

7. You are given a regression problem with dataset $x_i \in \mathbb{R}$ and $y_i \in \mathbb{R}$, where $i = 1, 2, \ldots, m$. Denote the prediction of the model on x_i as f_i. The cost function you optimize is the following

$$J = \frac{1}{m} \sum_{i=1}^{m} \mathcal{L}(f_i, y_i)$$

You have to decide on the following two loss functions

$$L_1(f_i, y_i) = |y_i - f_i|$$
$$L_2(f_i, y_i) = (y_i - f_i)^2$$

(a) Draw $\mathcal{L}_1(x, 0)$ and $\mathcal{L}_2(x, 0)$ versus $x \in \mathbb{R}$ on the same plot. [2 Marks]
(b) What is an outlier data sample in this context? Based on the plots, which loss works better if you know that there are a large number of outliers in your training data? [1+3 Marks] (Hint: Contributions of outliers to gradient calculations should be as small as possible)

8. Consider the CBoW configuration of the word2vec algorithm. During the training phase consider the following: vocabulary size is W, input (context) vector for each word x is represented by u_x, output (center) vector for word o is represented by v_o, the representation for words is of d dimensions, and the wondow size for context is k on each side.
(a) What are the dimensions of the of the input, hidden and output layer activations? [2 Marks]
(b) Clearly describe how forward pass operation happens in terms of the matrix operations. (how the context is represented, and how the hidden layer weights operate on it, etc.) [2 marks]
(c) How would you use the learned word embeddings to solve an analogy problem? (you need to predict the missing word here, $a : b ::? : d$) [2 Marks]

9. What is regularization in the context of deep learning, what is its significance? Clearly describe at least four different ways of regularizing a DNN. [2+4 Marks]

10. (a) Mention how the complexity of the mapping learned by the DNN increases with (i) depth of the model, and (ii) width of layers. [2 Marks]
(b) Clearly draw and describe the 'double-descent' risk curve in contrast with the conventional risk curve. (draw both curves and annotate them in detail) [4 Marks]

11. Batch Normalization is observed to be present in almost all the modern CNN architectures.
(a) Describe clearly the motivation behind Batch Normalization? [2 Marks]
(b) How is it realized during the training of a DNN? [2 Marks]
(c) How does it operate during the evaluation phase? [2 Marks]

12. (a) For small batch sizes, the number of iterations required for convergence decreases as the batch size increases. Why is that? [1 Mark]

(b) For large batch sizes, the number of iterations does not change much as the batch size is increased. Why is that? [1 Mark]

13. Consider encoder decoder model for machine translation.
(a) Using LSTM blocks as single elements, draw depth-1 recurrent neural network for language translation. Show sample input $(x_1, x_2, \ldots, x_n$ and corresponding output $(y_1, y_2, \ldots, y_m)$ words. [2 Marks]

(b) What is attention mechanism? Briefly describe how the encoder decoder system described above can inculcate attention mechanism. [2 Marks]