IIIT Vadodara
Winter 2021-22
(BTech) CS/IT308 Machine Learning
End-semester examination (REMOTE)
Instruction: Open self-handwritten notes and class slides
All questions are compulsory and carry equal marks.

Q. 1: In the course, we started with naive Bayes classifier and discussed upto a few contemporary machine learning models including both discriminative and generative models. We realized that neural network models are typically developed for a targeted application in a specific domain. Three critical domains include (i) defense & military, (ii) remote sensing, and (iii) medical & healthcare. Consider you have been hired to develop a suitable machine learning model for an application in one of these domains. However, as a machine learning engineer, one need to think and take care about certain items before developing model, during the model development, as well as robustly evaluating the developed model. Choose a preferred application in one of the three critical domains (as mentioned above) and develop a prototype neural model. You need to meticulously write down and provide technical details for each point at each of the three stages mentioned. Note that the selection of an application, a domain, and a model are not important; however, on the contrary, how in detail you cover all the small issues and discuss them to-the-point are important while answering this question.

Q. 2: Consider a linear support vector machine (SVM) and following training data in R^2,

$$\text{Class-1 } (y=-1)\text{: } \{\mathbf{x}_1=[1,1]^T, \mathbf{x}_2=[2,2]^T, \mathbf{x}_3=[2,0]^T\}$$

$$\text{Class-2 } (y=+1)\text{: } \{\mathbf{x}_4=[0,0]^T, \mathbf{x}_5=[1,0]^T, \mathbf{x}_6=[0,1]^T\}$$

(a) Plot the training points. Use '•' for Class-1 and 'o' for Class-2; (b) What are the support vectors?; (c) write the expression (by inspection) for the optimal hyperplane; (d) find by inspection, the weight vector **W** for the optimal hyperplane, and the optimal margin Δ; (e) Construct the solution framework in the dual space by constructing the Lagrange undetermined multipliers, α_i, i.e., need to provide the construction (linear system to be solved) and not required to solve. Hint for part (e): $\mathcal{L}_D(\alpha)=\sum_{i=1}^n \alpha_i - \frac{1}{2}\sum_{i,j}^n \alpha_i\alpha_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j$, such that $\sum_{i=1}^n \alpha_i y_i = 0$ for $\alpha_i \geq 0$.

Q. 3: Let us consider exclusive-OR (XOR) problem that belongs to nonlinearly separable category. We would like to solve it using nonlinear SVM. To this end, consider two-class classification problem with following training data in R^2:

$$\text{Class-1: } \{\mathbf{k}_1 = [1, 1]^T, \mathbf{k}_3 = [-1, -1]^T\}$$

$$\text{Class-2: } \{\mathbf{k}_2 = [1, -1]^T, \mathbf{k}_4 = [-1, 1]^T\}$$

(a) Plot the training points. Use '•' for Class-1 and 'o' for Class-2; (b) Now, we define a function $\phi : R^2 \rightarrow R^6$ so that we have following expansion: $1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, x_1^2, x_2^2$, where they can be linearly separated. Plot a 2-D projection of the mapped data points. (c) what are the support vectors? (d) By inspection, found the optimal hyperplane in the high dimensional space; (e) find, by inspection, the margin.

Q. 4: Referring to artificial neural networks (ANNs): (a) show that if the activation function of the hidden units is linear, then a three-layer ANN is equivalent to a two-layer ANN; (b) using the result of (a), explain why the three-layer ANN with linear hidden units cannot solve a nonl[illegible] problem such as XOR.

Q. 5: Consider a standard three-layer backpropagation network with d inpu[illegible]ts, n_H hidden units, c output units, and bias. (a) How many weights are in the network? (b) Consider the symmetry in the value of the weights. In particular, show that if the sign is flipped on every weight, the network function is unaltered.

Q. 6: Referring to the class notes and notations therein, the steepest-descent neural network training algorithm has now been modified to *batch-mode* processing. Show that steepest-descent weight update based upon ϵ_{avg} can be written as

$$\mu\Delta w_{l,j,i} = -\mu\frac{\partial\epsilon_{avg}}{\partial w_{l,j,i}} = -\frac{\mu}{N}\sum_{n=1}^{N} e_j(n)\frac{\partial e_j(n)}{\partial w_{l,j,i}} \qquad (1)$$

Q. 7: Define and differentiate discriminative neural models and generative neural models.

Q. 8: Compare and contrast variational autoencoder (VAE) and generative-adversarial network (GAN).

*** End of question paper ***