C2- Review Test; 20 April, 2022

Note: Attempt all questions. Submit hand written or Typed solution in a single file.

1. Support Vector Machine (SVM) can be viewed as minimizing hinge loss:

$$min_{a,b} \sum_{i=1}^{N} L_H(y,t) + \frac{1}{2\gamma} \|W\|_2^2$$

where hinge loss is defined as:

$$L_H(y,t) = max(0, 1 - ty)$$

.

   (a) TRUE/FALSE: If the total hinge loss is zero, then every training example must be classified correctly. Justify your answer. [1marks]

   (b) TRUE/FALSE: If the data set is linearly separable, then the optimal soft-margin SVM weights (according to above objective) must classify every training example correctly. Justify your answer.[1marks]

   (c) Suppose we replace the hinge loss with the following:

$$L_H(y,t) = max(0, ty)$$

   and otherwise keep the soft-margin SVM objective the same. What would go wrong? [1marks]

2. Consider one layer of a multilayer perceptron (MLP), whose computations are defined as

$$z_i = \sum_j w_{ij} h_j + b_i$$

$$y_i = \phi(z_i)$$

, where ϕ is a nonlinear activation function, h_j denotes the input to this layer (i.e. the previous layer's hidden units), and y_i denotes the output of this layer.

   (a) Give the backpropagation rules for $\bar{z_i}, \bar{h_j}$ and $\bar{w_{ij}}$ in terms of the error signal $\bar{y_i}$. [2 marks].

   (b) What happen if we will not use activation function ϕ in MLP? [1 Marks].

(c) How we define error function for 3 class problem? [1marks]

(d) What is the difference in MLP for regression and classification problem? [1marks]

3. Using maximum variance concept derive model for principle component analysis.[2 marks]

4. Write a program for 3-2-2 (3 nodes in input layes, 2 nodes in hidden layer and 2 node in output layer) ANN. You may use numpy and panda only. [3 marks].

5. What is the difference between Logistic regression and ANN? [2marks]