

Subject: 21DS602/21CS644

Assignment: 02

Distribution Date: 10-12-2021

Submission Due: 22-12-2021

Notes:

- This assignment is an optional work. It contributes towards the lab component of assessment
- The Max marks for the assignment is 32; this shall be reduced to 8 marks in final scoring
- The assignments are to be solved with writings. No computational work is needed.
- You may type the content or alternately write on paper. Submit the Assignment as scanned PDF or DOCX/PDF. Use figures & diagrams to illustrate your point.
- Please submit only a single file for the entire assignment.

Questions:

Q1. Explain Precision and Recall from the confusion matrix for binary and multi-class scenario. (4)

Q2. Explain the beta variation in F score evaluation. Explain the relation it has with the cost of misclassification. (4)

Q3. Explain with reasoning the scenario of underfitting and overfitting in Decision Trees. (4)

Q4. Find the values of the weights (W_0, W_1, W_2) after the first epoch of running through the OR gate truth table. Assume the initial conditions are (W_0, W_1, W_2) = (-2, 0.75, 0.63) and learning rate (α) = 0.1. (Round off the answers to 2 places after decimal, if exceeds 2 places.) Use Step activation function ($y = 1$ if $x \geq 0$; else $y = 0$). Use weight update formula for W_k is as follows: Weight change value, $\Delta W_k = \alpha * |\text{target} - \text{observation}| * \text{input}$.
 $W_k \leftarrow W_k + \Delta W_k$ (4)

Q5. Trace a Decision Tree based on the below provided data. Use Information gain & Gini index separately to trace two trees. (4)

Height	Weight	Hair Length	Sex (Label)
High	High	Long	M
Low	Average	Short	F
Med	Average	Short	M
Med	Low	Short	F
Low	Low	Long	M
High	Average	Long	F
Low	High	Long	F

Q6. A line for linear regression is represented by the formula $Y = mX + c$. For the given data values calculate the parameters m & c . Please round the answer values to first decimal point (Ex: 1.32 becomes 1.3 or 0.29 become 0.3).

Data = {(1, 1.75), (2, 2.14), (2.5, 2.47), (3, 2.38), (3.5, 3.55), (4, 2.72), (5, 2.91), (4.5, 4.2)} (4)

Q7. Following table contains the support vectors and weights. Find the class value for the vector ($A = [7 \ 3]$) using the formula provided below. (4)

Intercept (b) = -6.8.

Class value for test vector $X = \text{Sign}(\sum (y_i * \alpha_i * \text{DOT_PRODUCT}(\text{support vector}, X)) + b)$

Supprt Vectors	$y_i * \alpha_i$	Class Values
[1.7 0.5]	-0.4485	-1
[1.9 0.2]	-1	-1
[1.9 0.4]	-1	-1
[3.3 1]	0.436651	1
[5.1 1.6]	0.160646	1
[3 1.1]	1	1

Q8. Explain the role played by the kernel function in an SVM (illustrate with example). Explain the manners (design & decision making) in which a multi-class classification may be handled with SVM. (2+2)