

End Semester Examination

Course Code : EE871 Course Title : **Machine Learning**
 Credits : 3-1-2 : 5 Instructor : Dr. Jora M. Gonda
 Weight : 20% Marks : $5 \times 7 = 35$

NOTE: Answer all questions, each carry 5 marks.

1. Suppose we have a drawer full of dice, each of which has either 4, 6, 8, 12, or 20 sides. We do not know how many of each type are in the drawer. A die is picked at random from the drawer and rolled 5 times. The results in order are 4, 2, 4, 7, and 5. Given the priors: $H_4 = 0.20$, $H_6 = 0.25$, $H_8 = 0.20$, $H_{12} = 0.20$, and $H_{20} = 0.15$, estimate the possible die that was pulled from the drawer, and its probability.

2. Given the table:

(i) $\underline{x}(n)$	1	2	3	4	5 = I	$y(n)$ Liking
	shape	crust-size	crust-shade	fill-size	fill-shade	
(n)						
1	square	thick	grey	thin	white	like
2	triang	medi	green	medi	red	dislike
3	square	thin	red	thick	white	like
4	rect	thick	grey	medi	red	dislike
5	triang	thin	grey	thin	red	like
6	square	thin	grey	thin	dark	dislike
7	triang	medi	green	medi	white	like
8	square	thin	grey	thick	red	like
9	round	thick	grey	medi	dark	like
10	round	thick	red	thin	blue	like
11	triang	thick	yellow	thin	white	like
12	round	thin	blue	medi	dark	dislike

and the test input $\underline{x} = [\text{round, thick, grey, medi, dark}]$, determine the posteriors, $\Pr(\text{like}|\underline{x})$.

3. Given the results of a validation test on a model

	Predicted		
Actual	0	1	Total
0	217		263
1	26		252
Total	243	272	515

Fill up the empty cells and then compute the following Performance metrics:

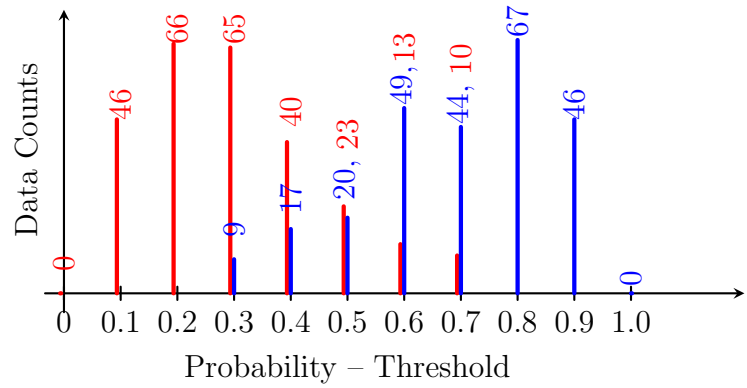
- i) Sensitivity ii) F1-Score.

4. List the indices used for deciding optimal threshold in classification.

5. Consider the following:

Given the data for two classes 'RED' and 'BLUE' as shown in the adjacent figure, determine the confusion matrix for $p \geq 0.7$ and enter in the matrix as follows.

$p \geq 0.7$	Predicted		
Actual	RED	BLUE	Total
RED			
BLUE			
Total			



6. Given a set of data as follows, for learning the concept, PlayTennis using Decision-Trees.

Day	Outlook	Temperature	Humidity	Wind	Play
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Rain	Mild	Normal	Weak	Yes
D10	Sunny	Cool	Normal	Strong	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Rain	Hot	High	Strong	Yes
D13	Overcast	Mild	Normal	Weak	Yes
D14	Overcast	Mild	High	Strong	No

Calculate the Entropies to decide the root node and find out the best Attribute to split on at the root.

7. Aircrew escape systems are powered by a solid propellant. The burning rate of this propellant is an important product characteristic. Specifications require that the mean burning rate must be 50 centimeters per second. We know that the standard deviation of burning rate is $\sigma = 2$ centimeters per second. The experimenter decides to specify a Type-I error probability or significance level of $\alpha = 0.05$ and selects a random sample of $N = 25$ and obtains a sample average burning rate of $\bar{x} = 51.3$ centimeters per second. What conclusion can be drawn and why?