



UNIVERSITY WITH A PURPOSE

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES

End Semester Examination, December 2021

Course: Pattern and Anomaly Detection
Program: B. Tech. CSE AIML
Course Code: CSAI 3007

Semester: V
Duration: 03 hrs.
Max. Marks: 100

Instructions:

- Read the questions carefully
- Answer each question in prescribed format and length.
- Formula sheet or data sheet will not be provided
- Internal choices are given in questions number 6 and 10. Attempt any one of the choices.

SECTION A
Short answer type (up to 120 words)

(5QX4M = 20 Marks)

		Marks	COs
Q1	Briefly describe how the process of cross-validation in machine learning based pattern recognition works. You might need the help of training, validation, and testing samples	4	CO1
Q2	Are Bernoulli distributions parametric and are used for continuous variables? Answer yes or no. Also, support your answer with evidence.	4	CO1
Q3	In least squares solution for weights in linear models for regression, what do we minimize? Explain with formula.	4	CO2
Q4	When do we say, a machine learning based classification model is “linear”? Discuss briefly	4	CO3
Q5	Explain the formula for model coefficients (weights) of a linear regression model obtained using maximum likelihood with least squares estimation method.	4	CO4

SECTION B
Medium answer type (up to 350 words)

(4QX10M = 40 Marks)

		Marks	COs
Q6	Define overfitting in machine learning. Discuss the concept of regularization and its use in machine learning to avoid overfitting? Enumerate different regularization types. (2+5+3)	10	CO1
	OR Discuss the concepts of hyperparameter tuning in context to machine learning. Review the different hyperparameter-space search strategies. (4+6)		

Q7	Name 5 regression models. Discuss the relationship between input-output and model parameters-output in these models. (3+7)	10	CO5
Q8	Define the term 'outliers'. State the simplest method to detect outliers in data. Discuss the local outlier factor (LOF) method of outlier detection. (2+2+6)	10	CO3
Q9	Define the term 'kernel' in context to machine learning based pattern recognition. State the key concept in kernel based models. Outline the key differences between kernel models and sparse kernel models. Give two examples of kernel models and sparse kernel models. (3+4+3)	10	CO2, CO4

SECTION C
Long answer type (upto 700 words)

(2QX20M = 40 Marks)

		Marks	COs
Q10	<p>In context to neural networks (NNs) applications to real-world problems</p> <ol style="list-style-type: none"> Highlight its key advantage over other models i.e. what makes it so special? (2) Discuss the role, an activation function plays in NN models. Can we call them basis function as well? (4) How do neural networks learn? Discuss briefly the error backpropagation algorithm briefly with appropriate formulas. (8) Define linearly separable datasets. Can multilayer perceptron (type of NNs) separate non-linearly separable datasets? Support your answer. (6) <p style="text-align: center;">OR</p> <p>Discuss on the following neural network training algorithms and derive how weights (model coefficients/parameters) are computed in each case with the help of formulas.</p> <ol style="list-style-type: none"> Perceptron algorithm (4) Pocket algorithm (4) Gradient descent (4) Stochastic gradient descent (4) Error backpropagation (4) 	20	CO1, CO3, CO5
Q11	<p>Suppose you wish to build a Bayesian regression model. Discuss the advantage of building Bayesian regression model. Suppose you considered the following</p> <p>A likelihood distribution of the form</p> $p(\mathbf{t} \mathbf{X}, \mathbf{w}, \beta) = \prod_{n=1}^N \mathcal{N}(t_n \mathbf{w}^T \phi(\mathbf{x}_n), \beta^{-1}).$ <p>and a prior distribution of the form</p> $p(\mathbf{w}) = \mathcal{N}(\mathbf{w} \mathbf{m}_0, \mathbf{S}_0).$ <p>Using this likelihood and prior, you obtained the following posterior distribution</p> $p(\mathbf{w} \mathbf{t}) = \mathcal{N}(\mathbf{w} \mathbf{m}_N, \mathbf{S}_N)$ <p>All variables have usual meaning. Derive the formulas for posterior mean (\mathbf{m}_N) and covariance (\mathbf{S}_N). Also discuss how the posterior mean and covariance changes during training. Hint:- This is Bayesian learning, therefore learning will be one sample at a time. (5+10+5)</p>	20	CO2, CO4