IAI, TCG-CREST December 19 2022

Machine Learning

Final Semester Examination

Ph.D. Program Session 2022-2023

Time Duration: 3 hours. *Full marks: 60.*

Answer any **6** of the following 10 questions.

1. What are the feature scaling approaches of min-max standardization and mean-standard-deviation normalization. Define metric. What are kernel functions, and what is Mercer's condition for kernel functions to be a valid kernel. (4+3+3)

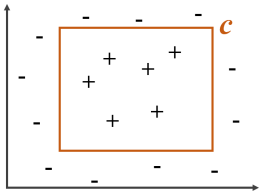
2. What are the Ridge Regression and Lasso Regression model? Derive update expressions for the Ridge Regression model parameters. Let S be a convex subset of $\mathbb{R}^n$, then prove that $f : \mathbb{R}^n \to \mathbb{R}$ is convex iff for any $\mathbf{x}, \mathbf{y} \in S$ we have $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x})$. (3+2+5)

3. What are limitations of the softmax regression approach that motivated the development of Multi-Layered Perceptrons (MLPs)? Describe the feedforward operation in a single hidden layer MLP leading to calculation of a loss function. Derive backpropagation update expressions for the single hidden layer MLP parameters. (2+3+5)

4. Describe how a Recurrent Neural Network (RNN) cell is defined. Describe the forward propagation operation of an RNN cell. When backward propagation through time is done for RNNs, what are the challenges of exploding gradients and vanishing gradients? (5+2+3)

5. For classification validation, what is a Confusion Matrix, and what are Accuracy and Error Rates? Define Precision, Recall, and F-measure. In the context of binary classification, define True Positives, True Negatives, False Positives, and False Negatives. (3+3+4)

6. In the $\mathbb{R}^2$ space, let the concept class be a set of rectangles with length and width parallel to the two axes. An example of a specific concept is shown in the figure below.



   Derive PAC bounds on the number of training instances required to learn a hypothesis with error ϵ.

   Define Shattering and VC dimensions. For data instances in $\mathbb{R}^2$, what is the VC dimension of the hypothesis set of straight lines? (4+2+4)

7. Formulate a maximum hyperplane objective and derive the constrained quadratic optimization objective function for Support Vector Machines (SVMs). Show how the objective function of SVMs with soft constraints can be reformulated to a constrained hinge loss objective function. (5+5)

8. For the AdaBoost classifier, define the error at the m-th step, and derive the expressions to learn instance weights and the weight of the m-th base learner. State the overall Adaboost algorithm. (6+4)

9. Describe the k-Means++ procedure for cluster center initialization. From the kernel k-Means objective function, derive update expressions for the cluster memberships. In the context of hierarchical clustering, define single linkage, average linkage, and complete linkage. (4+3+3)

10. Define Entropy. Show how using Information Gain, a Decision Tree can be learnt to predict the target variable of PlayTennis for the training data in Table 1. (2+8)

Day	**Outlook**	**Temperature**	**Humidity**	**Wind**	**PlayTennis**
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	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Table 1: Training data to predict the target variable PlayTennis