CMR Engineering College :: Hyderabad

II - M.Tech - I Sem End Exam (Reg) - Jan 2026

AI and Machine Learning.

(VLSI - SD) (R22EC5730IPE)

PART - A

1. a. * Entropy
* Information gain
* Gain Ratio
* Gini Index
* Chi_Square

b. Ranking is the process of ordering instantes or items based on a score, relevance, or probability Predicted by a model.

c. Dimensionality reduction is the process of reducing the numbes of input features while Preserving important information.
⇒ Reduces Computational Cost
⇒ Avoids overfitting
⇒ Improves model Performance

d) d. To transform correlated variables into uncorrelated principal components while retaining maximum variance

e) Statistical learning is a framework that uses probability, statistics and optimization to learn patterns from data & make predictions

---

f) Bagging :-

It means BootStrap Aggregating is an ensemble method where multiple models are trained on different bootstrapped samples and their outputs are aggregated

---

g) XOR is not linearly seperable, and a single-layer perceptron can only solve linearly seperable problems.

---

h) Back propagation computes the error at o/p propagates it backword through layers, and updates weight using gradient descent.

---

i) Inference mechanism applies logical rules or probabilistic reasoning to derive conclusions from given facts or knowledge.

---

j)
* Centroid
* Bi Sector
* Mean of Maximum
* Smallest of maximum
* Largest of Maximum

2.
a. Yes. Machine learning cannot solve all problems

Justification:-

* Machine learning is a powerful technique that enables systems to learn patterns from data & make predictions or decisions. However it has several inherent limitations that prevent it from being a universal solution for all problems.

⇒ Requirements of large & quality of data
⇒ Inability to handle well-defined rule based problems efficiently
⇒ Lack of explainability & transparency
⇒ Poor Generalization outside training data
⇒ Dependence on human expertise
⇒ Ethical, Legal & social constraints

Machine learning is highly effective for pattern recognition, prediction & data driven decision making, it is not suitable for all types of problems.

---

b) Linear classification is a supervised machine learning technique used to classify data points by seperating them into different classes using a linear decision boundary. The classifier learns a linear function of the input features and assigns a class label based on the o/p of this function:

* It is represented as,

$$f(x) = w^T x + b$$

where,

$w \rightarrow$ weight vector

$x \rightarrow$ i/p feature vector.

$b \rightarrow$ bias

The sign / value of this function determines the class label.

* This method work effectively a linearly seperable that a 2D, 3D or hyperplane can seperate the classes.

* Eg perceptron, logistic regression & linear svm

* The main advanteges of linear classification are its simplicity, fast training & good interpretability. However it has limitations when handling non linear data, overlapping classes or complex decision boundaries.

---

3. a * It's a supervised learning problem in which an i/p sample is assigned to one of more than 2 possible classes. unlike binary classification, multiclass problems involve three or more mutually exclusive class labels.

* In multiclass classification, the model learns a decision function that seperates multiple classes simultaneously.

⇒ OVR → A seperate binary classifier is trained for each class against all other classes

⇒ OVO → A classifier is trained for every pair of classes.

⇒ native multiclass models – Algorithms like Softmax regression, Decision trees & neural networks handle multiclass problems directly

⇒ probabilities represent the model's confidence that a given input belongs to each class. A widely used method to compute class probabilities is the softmax function,

$$P(y=k/x) = \frac{e^{z_k}}{\sum_{i=1}^{c} e^{z_i}}$$

3. b. * It is a supervised machine learning task in which each i/p instance is assigned to one of 2 possible classes, commonly labeled as Positive/Negative, Yes/No, 1/0 etc.

* In binary classification, the model learns a decision function that seperates the two classes based on i/p features. This function produces a score for each instance, usually a real valued number indicating how strongly the instance belongs to the positive class.

* Scoring refers to assigning this numerical value to each instance. A higher score indicates a higher likelihood of belonging to the positive class. By applying a predefined threshold, the score is converted into a final class label.

* Ranking involves ordering instances based on their scores rather than assigning hard class labels. Instances with higher scores are ranked higher.

(A) It is a type of a machine learning in which the model is trained using unlabeled data i.e no predefined O/P or class label is provided

Types:

(i) clustering - A groups of similar data points into clusters based on distance or similarity measures. Objects within same cluster are more similar to each other than to those in other clusters. Eg K-means clustering

(ii) Association Rule learning:-

This technique identifies interesting relationships & dependencies among variables in large datasets. It is used in market based Analysis.

(iii) Dimensionality Reduction:-

It reduces the number of features while preserving important information. It helps in data visualization, noise reduction & improving model efficiency

(iv) Anamoly Detection:- This method identifies data points that significantly deviate from normal pattern. It is useful for fraud detection Network security & fault diagnosis.

(5) PCA → It is an unsupervised statistical technique used to reduce the dimensionality of a data set while preserving as much variance as possible. It transform the original correlated features into a smaller set of uncorrelated variables called principle components. (4)

Use of PCA :-

* Data standardization - I/P data is first normalized to have zero mean & unit variance to ensure that all features contribute equally
* Covariance matrix computation - Used to measure the relationships b/w different features.
* Eigen values and Eigen vectors :-
Eigenvectors of the covariance matrix determine the directions of maximum variance while eigen values indicate the amount of variance captured by each direction
* Selection of principle components - Eigenvectors corresponding to the largest eigen values are selected as principal components. Only top k components are retained.
* Data projection - original data projected onto the principle components, resulting in a lower dimensional representation

6. Statistical learning theory provides a rigorous mathematical framework for understanding learning generalization & model selection in machine learning. This shows how a learning algorithm can infer an underlying function from finite samples drawn from an unknown probability distribution.

* The objective function is used to minimise the expected risk,

$$R(f) = E_{(x,y)}[L(y, f(x))]$$

where,

$L(\cdot)$ is a loss function. Since the distribution is unknown, learning algorithms minimise the empirical risk computed over training samples. However, minimizing empirical risk alone may lead to overfitting.

* Bias-variance-trade off explains the compromise between approximation error & estimation error. SVM, where maximizing the margin improves generalization

---

7 Random forest is an ensemble learning method that builds a large number of decision trees & combines their predictions to improve accuracy and generalization. It is based on the principle of bagging with additional randomness

* Steps in random forest :-

1. Bootstrap Sampling - Multiple datasets are created by random sampling with replacement from the original training data

2. Tree construction with feature randomness :-

For each decision tree, at every split, only a random subset of features is considered instead of all features.

3. Independent tree training

Each tree is grown independently & usually to full depth without pruning

4. Aggregation Of Predictions — Classifications Regression

This randomness reduces correlation among trees leading to lower variance & better generalization

S.no.	Bagging	Boosting	Random forest
1.	Reduce variance	Reduce bias & variance	Reduce variance + decorrelation
2.	Bootstrap Sampling for data	sequential reweighting	Bootstrap Sampling
3.	Independent tree dependency.	Sequential & dependency	Independent
4.	All selected features are used.	All selected features are used	Random features & Subset.
5.	Moderate overfitting	Can overfit over noisy data	Strong

8. It is a computational model inspired by the structure of the human brain. It consists of interconnected processing units called artificial neurons, organized into I/P, hidden & O/P layers.

* Each neuron receives I/P sig, multiplies them by corresponding weights, adds bias & computes a weighted sum

$$z = \sum w_i x_i + b.$$

* This value is passed through an activation function to introduce non linearity & generate the neuron's O/P

* During training, the ANN uses labeled data & loss function to measure prediction error. It is combined with gradient descent, updates the weights by propagating the error backward through the network to minimize the loss.

S.No	Artificial Neural Network	Biological Neural Network
1.	Artificial Neuron	Biological neuron.
2	Signal types is numerical values	Electro chemical signals
3.	Follows back propagation & optimization	Synaptic plasticity.
4.	Fault tolerence is limited	very high.
5.	Signal processing speed is very fast.	Slower but massively Parallel.
6.	High Power consumption	Extremely energy Efficient

9. a. * The multilayer neural networks, hidden layers (6) introduce non linear transformations that enable the network to learn complex i/p & o/p relationship
* Back propagation is required to efficiently train such networks by systematically updating weights in all layers.
* Back propagation works by computing the error gradient of the loss function with respect to each weight using the chain rule of calculus. The error is first calculated at the op layer & then propagated backward through the hidden layers. This allows each neuron to receive information about its contribution to the overall error.
* without back propagation, only single layer networks could be trained effectively, limiting the model to linearly seperable problems.
* Back propagation enables end to end learning, allowing deep networks to adjust internal representations & learn hierarchical features.
* By combining back propagation with optimization methods such as gradient descent, the network minimizes the loss function iteratively, ensuring convergence & generalization.

9.
b. Radial basis function (RBF) - It is a type of feed forward Neural Networks that uses radial basis function as activation functions in the hidden layer. It is commonly used for function approximation, classification & pattern recognition due to its fast training & strong interpolation capability.

* RBF N/w has — i/p layer, hidden layer, o/p layer

* RBF neuron = $\phi(x) = \exp\left(-\frac{\|x - c_i\|^2}{2\sigma_i^2}\right)$

where,

$c_i \rightarrow$ center & σ_i - width of the basis function

* Each hidden neuron responds strongly only to inputs close to its center.

* The network o/p is a weighted sum of these localized responses.

* Ex: Non linear function approximation

$$y = \sin(x)$$

⇒ i/p layer = x

⇒ hidden layer = Gaussian RBF centered at different values of x

⇒ o/p layer - weighted sum approximates the sine curve

* Advantages - Fast & Stable training, Good performance, Simple structure

* Limitations - choice of centers & width is critical & poor scalability for very high dimensional data

10. * Knowledge representation is the process of encoding information about the real world in a form that a computer system, such as AI, can utilize to solve complex tasks such as reasoning decision making & Problem solving. It is essential because raw data or facts are insufficient for intelligent behavior, the system must organize knowledge to enable effective reasoning

*Types/methods.

1. Logical Representation
2. Semantic Networks
3. Frames
4. Rule based systems
5. Production systems

*Inference mechanism:-

It is a computation process that derives new knowledge or conclusions from existing knowledge using logical reasoning. It enables AI systems to solve problems, solve queries or make decisions.

⇒ Types:

(i) Forward Chaining
(ii) Backward Chaining
(iii) Rule based Reasoning
(iv) Probabilistic Reasoning
(v) FUZZY Reasoning.

11. Defuzzification is the process of converting the fuzzy o/p of a fuzzy inference system into a crisp numerical value that can be used for control or decision making. This step is essential because fuzzy reasoning produces fuzzy sets, which cannot directly drive real-world problems.

* Defuzzification comes after,
  (i) Fuzzification
  (ii) Rule evaluation/ inference
  (iii) Aggregation of fuzzy o/p's.

* Common Defuzzification methods:-

1. Centroid method.

$$z^* = \frac{\sum x_i \cdot \mu(x_i)}{\sum \mu(x_i)}$$

2. Bisector of Area - Finds the value that divides the total area under the fuzzy set into 2 equal halves.

3. Mean of maxima - Averages all o/p values that have the maximum membership degree

4. Smallest / Laggest of maximum:-

* Working logic:

⇒ The hidden layer implements membership functions and fuzzy rules.

⇒ The inference mechanism aggregates fuzzy o/p.

⇒ The o/p layer performs defuzzification to produce the crisp o/p. (8)

* Applications:

⇒ Control Systems
⇒ Robotics
⇒ Consumer electronics
⇒ Automotive Industry