BIO542: Machine Learning for Biomedical Applications

(18th November 2024, Quiz3)

Maximum Marks: 20 **Duration: 20 Minutes**

Name: _____ **Admission No:** _____

Instructions: Attempt all questions; each question carries one mark.

1. Which of the following technique can be used for data imputation?
   a) **K-Nearest Neighbors** b) Principal Component Analysis
   c) Normalization d) Log transformation

2. Which method is used to handle outliers in data?
   a) Data encoding b) **Z-score**
   c) Feature scaling d) Log transformation

3. What does UMAP stand for in feature engineering?
   a) **Uniform Manifold Approximation and Projection**
   b) Universal Model Analysis and Projection
   c) Univariate Model Approximation and Processing
   d) Unified Multi-dimensional Approximation and Projection

4. Which scaling technique transforms values to a range of 0 to 1?
   a) Standardization b) L2 normalization
   c) **Min-Max Normalization** d) Log transformation

5. Which of the following method is not used for feature selection
   a) Embedded method b) Filter method
   c) **Alignment method** d) Wrapper method

6. Which method is used for feature selection if input and output are numerical variables
   a) **Spearman correlation** b) ANOVA
   c) Kendall's d) Alignment

7. Which type of RNA is primarily translated to polypeptide/protein.
   a) **mRNA** b) tRNA
   c) rRNA d) miRNA

8. 8. In FASTAQ format line 4 contain information about
   a) Sequence identifier b) **Quality scores**
   c) Raw sequence d) Nothing

9. Which of the following techniques can be used to cellular heterogeneity?
   a) Microarray b) **Single-cell Sequencing**
   c) Bulk-RNA sequencing d) Whole genome sequencing

10. Which color dye is used in Stanford/cDNA microarray technique
    a) Blue
    b) **Green**
    c) Yellow
    d) Orange

11. Which server developed for measuring progression of skin cutaneous melanoma?
    a) CancerPred
    b) CancerDR
    c) SKCMhrp
    d) **CancerSPP**

12. Which of the following is not part of RNA-Seq technology
    a) RNA isolation
    b) mRNA enrichment
    c) **Labelling of cDNA**
    d) Adapter ligation

13. Which of the following unit is not measure read counts in RNA-Seq
    a) RPKM
    b) **RTPM**
    c) FPKM
    d) TPM

14. Which of the following is a database of gene expression profiles?
    a) **ArrayExpress**
    b) GeneExpress
    c) GeneBank
    d) SwissProt

15. Which technique is commonly associated with bagging?
    a) Gradient Boosting
    b) **Random Forest**
    c) AdaBoost
    d) SVM

16. Gini Index is used to measure _____ in decision tree?
    a) Correlation values
    b) **Impurity in a dataset**
    c) MCC
    d) Prediction accuracy

17. Which algorithm is used for reducing overfitting in decision tree by pruning trees?
    a) Random Forests
    b) AdaBoost
    c) **CART**
    d) XGBoost

18. Which algorithm was developed as a successor to ID3 and addresses its limitations?
    a) CART
    b) Random Forests
    c) **C4.5**
    d) XGBoost

19. Which ensemble learning technique, sequentially train a series of weak models to correct the errors made by its predecessors?
    a) **Boosting technique**
    b) Bagging technique
    c) Random forest
    d) Neural network

20. If a bag of contain 4 red and 4 blue balls then following will be gini index
    a) 0
    b) 1
    c) **0.5**
    c) 0.25