Worksheet set 2

STATISTIC WORKSHEET

1. d) Expected

2. c) Frequencies

3. c) 6

4. b) Chisquared distribution

5. d) Poisson Distribution

6. b) Hypothesis

7. a) Null Hypothesis

8. a) Two-tailed

9. b) Research Hypothesis

10. a) np

MACHINE LEARNING

ANS.1

- R-squared is commonly used for linear models, while RSS directly quantifies model error.
- For linear models, R-squared is preferred due to its ease of interpretation.
- However, in non-linear models, RSS remains relevant, especially while using non-linear least square (NLS) estimation.

ANS.2

TSS (Total Sum of Squares):TSS measures the total variability in the dependent variable (response) around its mean.It represents the sum of squared differences between each data point and the overall mean of the response.

Mathematically: TSS=i=1∑n(yi−y¯)2

, where (yi) is the observed value and (\bar{y}) is all known observations' average.

ESS (Explained Sum of Squares):ESS quantifies the variability explained by a regression model.It measures the sum of squared deviations of the predicted values from the mean of the response.ESS represents how well this model fits data.

Mathematically: ESS=i=1∑n(y^i−y¯)2

, where (\hat{y}_i) is an estimated value.

RSS (Residual Sum of Squares):RSS captures the unexplained variability in the data.It represents the sum of squared differences between observed values and predicted values, called residuals'.

Mathematically: RSS=i=1∑n(yi−y^i)2

, where (y_i) is an observed value and (\hat{y}_i) is anticipated value.

Relation between TSS,ESS& RSS.

TSS=ESS+RSS.

ANS.3

Regularization in Machine Learning:

- Regularization techniques prevent overfitting by adding penalties to the loss function.
- They control model complexity and reduce variance.
- Common methods include L1 (Lasso) and L2 (Ridge) regularization, which shrink coefficient estimates toward zero.
- Regularization helps generalize better to unseen data.

ANS.4

Gini Impurity Index:

- Gini impurity measures the impurity or disorder in a dataset.
- It ranges from 0 (pure, all samples belong to one class) to 0.5 (maximum impurity, equal distribution across classes).
- Used in decision trees for splitting nodes based on feature impurity.
-

ANS.5

Unregularized Decision Trees and Overfitting:

- Unregularized decision trees can grow deep and fit training data perfectly.
- They tend to overfit by capturing noise and specific patterns in the training set.
- Pruning or using ensemble techniques (like random forests) helps mitigate overfitting.

ANS.6

Ensemble Techniques in Machine Learning:

Ensemble learning combines multiple models to improve predictive performance.

Techniques include:

Bagging: Creates an ensemble by training multiple models independently and averaging their predictions (e.g., Random Forest).

Boosting: Sequentially builds an ensemble by correcting errors made by previous models (e.g., Gradient Boosting, XGBoost).

ANS.7

Difference Between Bagging and Boosting:

Bagging: Parallel ensemble, reduces variance, uses bootstrapped samples, and averages predictions.

Boosting: Sequential ensemble, corrects errors, adapts to misclassified instances, and combines weak learners into a strong one.

ANS.8

Out-of-Bag (OOB) Error in Random Forests:

OOB error estimates model performance without the need for a separate validation set.It uses samples not included in the bootstrap sample for each tree.Useful for assessing Random Forest accuracy.

ANS.9

K-Fold Cross-Validation: Divides data into K folds, trains on K-1 folds, and validates on the remaining fold.Repeats K times, rotating the validation fold.Provides a robust estimate of model performance and helps tune hyperparameters.

ANS.10

Hyperparameter Tuning:

Hyperparameter tuning is the process of selecting optimal values for a machine learning model's hyperparameters. These hyperparameters control the learning process itself and are set before training begins. Unlike model parameters (weights and biases learned from data), hyperparameters are not directly learned from the training process.Hyperparameter tuning is the process of selecting optimal values for a machine learning model's hyperparameters. These hyperparameters control the learning process itself and are set before training begins. Unlike model parameters (weights and biases learned from data), hyperparameters are not directly learned from the training process.

Benifits:

- Adjusting hyperparameters (e.g., learning rate, regularization strength) to optimize model performance.
- Done via techniques like grid search or random search.
- Helps find the best configuration for a given problem.

ANS.11

Issues with Large Learning Rate in Gradient Descent:

- Large learning rates cause overshooting, leading to divergence.
- The model may fail to converge or oscillate around the optimal solution.
- Smaller learning rates are preferred for stability.

ANS.12

Logistic Regression for Non-Linear Data: Logistic Regression is primarily designed for linear classification tasks.It assumes a linear decision boundary in the feature space.When dealing with non-linear data, Logistic Regression may not perform well because it cannot capture complex relationships.For non-linear data, other models (such as SVMs with non-linear kernels, decision trees, or neural networks) are more suitable.

ANS.13

Differentiate between Adaboost and Gradient Boosting:

Both Adaboost and Gradient Boosting are ensemble learning techniques that combine multiple weak learners to create a stronger model.

Adaboost:

- Focuses on hard training examples (those with high error).
- Adapts sample weights to emphasize misclassified instances.
- Sequentially trains weak learners, adjusting weights at each step.
- Combines weak learners into a strong model.

Gradient Boosting:

- Works on the principle of stagewise addition.
- Trains multiple weak learners sequentially.
- Each weak learner corrects the residuals of the previous one.
- Uses shrinkage to avoid overfitting.
- Handles both classification and regression tasks.

ANS.14

Bias-Variance Trade-Off in Machine Learning:

The bias-variance tradeoff is a crucial concept in machine learning. It describes the delicate balance between two types of errors in a model:

Bias Error: This occurs when a model makes overly simplistic assumptions about the underlying data, leading to systematic inaccuracies. High bias can cause underfitting, where the model fails to capture complex patterns.

Variance Error: Variance reflects how much the model's predictions vary when trained on different subsets of the data. High variance can cause overfitting, where the model fits the training data too closely and performs poorly on unseen data.

Striking the right balance between bias and variance is essential for optimal model performance.

ANS.15

Linear Kernel:

The linear kernel is the simplest and most straightforward.It computes the dot

product between input data points in the original feature space.Useful for linearly separable data.The decision boundary is a straight line or hyperplane.Implemented in SVMs as LinearSVC when using a linear kernel.

Example: Separating classes with a straight line in 2D space.

RBF (Radial Basis Function) Kernel:

The RBF kernel is powerful and versatile.Maps data into a high-dimensional space using a Gaussian-like function.Combines multiple polynomial kernels to handle non-linearly separable data.Widely used in SVMs for classification.Requires tuning of parameters (C and gamma) for optimal performance.

Example: Separating classes with curved decision boundaries.

Polynomial Kernel:

The polynomial kernel represents similarity using polynomial functions.Maps data into a higher-dimensional space.Allows learning of non-linear models.Interaction features (combinations of input features) are considered.Popular in natural language processing (NLP) tasks.Degree parameter (d) specifies the polynomial order (e.g., quadratic, cubic).

Example: Separating classes using polynomial curves.