**Ans 1)** Stages of Hypothesis

Data Sources

PreProcess Data

Transform Data

**Ans 2)** Standard approach to supervised learning:

Classification

Regression

Anomaly detection

**Ans 3)** Training set- In machine learning, a training set is a dataset used to train a model. In training the model, specific features are picked out from the training set. These features are then incorporated into the model. Thereby, if the training set is labeled correctly, the model should be able to learn something from these features.

Test data- The test set is a dataset used to measure how well the model performs at making predictions on that test set. If the prediction scores for the test set are unreasonable, we’ll need to make some adjustments to our model and try again.

**Ans 4)** Ensemble Method: combining different Models to get the results

Bagging: Bagging stands for bootstrap aggregation. One way to reduce the variance of an estimate is to average together multiple estimates.

Boosting: Boosting refers to a family of algorithms that are able to convert weak learners to strong learners. The main principle of boosting is to fit a sequence of weak learners− models that are only slightly better than random guessing, such as small decision trees− to weighted versions of the data. More weight is given to examples that were misclassified by earlier rounds.

**Ans 5)** Overfitting happens when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data. This means that the noise or random fluctuations in the training data is picked up and learned as concepts by the model. The problem is that these concepts do not apply to new data and negatively impact the models ability to generalize