**SUDHARSHANA B**

**22CSR208**

**ASSIGNMENT 1:**

|  |  |  |  |
| --- | --- | --- | --- |
| STUDENT ID | STUDY HOURS | PRACTICE TESTS | EXAM SCORE |
| 1 | 5 | 2 | 80 |
| 2 | 3 | 1 | 65 |
| 3 | 7 | 3 | 90 |
| 4 | 4 | 2 | 75 |
| 5 | 6 | 3 | 85 |
|  |  |  |  |

**FEATURE: Individual measurable attributes or traits used as inputs to the model**

**Example: the features are study hours and practice tests**

**LABEL: The output variable that the model aims to forecast**

**Example: the label is exam score**

**PREDICTION: The output from the model after it has been trained**

**Example: study hour and practice test will predict an exam score**

**OUTLIER: A data point that deviates significantly from other observations**

**Example: if one student had an exam score of 100 while others are between 65-90 then this score will be considered an outlier**

**TEST DATA: The portion of the data used to assess the model's performance**

**Example: students 4 and 5 can be used as test data to evaluate the model trained on the other students**

**TRAINING DATA: The portion of the data used to train the model**

**Example: Student 1, 2, 3 can be used as training data**

**MODEL: A mathematical representation of the relationship between features and labels**

**Example: a linear regression model that predicts exam scores based on study hours and practice tests**

**VALIDATION DATA: A subset of the training data used to adjust hyperparameters and avoid overfitting**

**Example: using student 3's data to validate the model trained on students 1 and 2**

**HYPERPARAMETER: Parameters whose values are set before the learning process begins**

**Example: learning rate or the number of epochs**

**EPOCH: One complete pass through the entire training dataset**

**Example: if we iterate through all the students' data once during training it counts as one epoch**

**LOSS FUNCTION: A method to evaluate how well the model's predictions match the actual data**

**Example: mean squared error (mse) could be used to measure the difference between the predicted and actual exam scores**

**LEARNING RATE: A hyperparameter that controls how much to change the model in response to the estimated error each time the model weights are updated**

**Example: a learning rate of 0.01 means the model is updated slowly while a learning rate of 1 means the model is updated quickly**

**OVERFITTING: When a model learns the training data too well including noise and outliers leading to poor performance on new data**

**Example: model that perfectly predicts the training data but performs poorly on the test data**

**UNDERFITTING: When a model is too simple to capture the underlying pattern in the data leading to poor performance on both training and new data**

**Example: a model that performs poorly on both the training and test data**

**REGULARIZATION: Techniques used to reduce overfitting by penalizing complex models**

**Example: adding a penalty to the loss function for large coefficients in a linear regression model**

**CROSS-VALIDATION: A technique to evaluate the model's performance by dividing the data into several subsets and training/testing the model on different combinations of these subsets**

**Example: 5-fold cross-validation involves splitting the data into 5 parts, training the model on 4 parts and testing on the remaining part, repeating this process 5 times**

**FEATURE ENGINEERING: The process of creating new features or modifying existing ones to improve model performance**

**Example: creating a new feature total study time by multiplying study hours with practice tests**

**DIMENSIONALITY REDUCTION: Techniques to reduce the number of features while retaining important information**

**Example: principal component analysis (PCA) to reduce study hours and practice tests into a single combined feature**

**BIAS: The error introduced by approximating a real-world problem which may be complex by a much simpler model**

**Example: a high bias model may consistently predict exam scores far from the actual scores**

**VARIANCE: The error introduced by the model's sensitivity to small fluctuations in the training set**

**Example: a high variance model may predict exam scores very accurately for training data but poorly for test data**

**Top of Form**

**Bottom of Form**