Logistic Regression

Logistic Regression is a widely used statistical method for binary classification. Imagine you're in a scenario where you want to predict whether students will pass or fail based on the number of hours they study. That's where Logistic Regression steps in. Despite its name, it's not diving into regression; instead, it's your trusty guide for figuring out if something belongs to one of two categories.

$$P(Y=1) = \frac{1}{1+e^{-(\beta_0+\beta_1 X)}}$$

This sigmoidal function ensures outputs between 0 and 1, facilitating the interpretation as probabilities. The parameters β0 and β1 are iteratively adjusted during training to optimize the model's predictive capabilities.

Algorithm Explanation:

- Initialization:
  Commencing with initial coefficients, the algorithm initializes the optimization process.
- Hypothesis:
  The hypothesis function incorporates the logistic function, encapsulating the probabilistic nature of the classification task.
- Cost Function:
  The log-loss function, representing the disparity between predicted and actual values, serves as the guiding metric for optimization.

$$\text{Cost} = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log(\hat{y}^{(i)}) + (1-y^{(i)})\log(1-\hat{y}^{(i)})\right]$$

- Gradient Descent:
  Employing optimization techniques such as gradient descent, the algorithm minimizes the cost function, refining the coefficients for enhanced predictive accuracy.

$$\beta_j = \beta_j - \alpha\frac{\partial}{\partial\beta_j}\text{Cost}$$

- Predictions:
  Post-training, the learned coefficients allow the algorithm to predict the probability of an instance belonging to the positive class.

Evaluation Metrics:

1) Confusion Matrix:
   A foundational table detailing the accuracy of predictions, distinguishing true and false positives/negatives.

2) Accuracy:
The ratio of correctly predicted instances to the total, offering a holistic view of the model's correctness.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

3) Precision:
The precision metric discerns the proportion of correctly predicted positive instances among the predicted positives.

$$\text{Precision} = \frac{TP}{TP+FP}$$

4) Recall (Sensitivity):
Emphasizing the algorithm's ability to capture all relevant positive instances, recall quantifies true positives against actual positives.

$$\text{Recall} = \frac{TP}{TP+FN}$$

5) F1-Score:
The harmonic mean of precision and recall, providing a balanced assessment of the model's efficacy.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

6) ROC-AUC Curve:
The Receiver Operating Characteristic curve and its Area Under the Curve metric delineate the model's ability to trade sensitivity for specificity.

Decision Tree Regression

Decision Tree Regression is a non-linear regression algorithm used for both classification and regression tasks. It operates by recursively partitioning the dataset into subsets based on specific conditions, forming a tree-like structure of decisions.

The decision tree regression algorithm makes predictions by recursively splitting the dataset based on features and their corresponding split points. At each node, the algorithm selects the feature and split point that minimizes the variance of the target variable within each subset. The prediction for a new instance is the mean (or median) of the target values in the leaf node to which the instance belongs.

Algorithm Explanation:

- Tree Construction:
  Begin with the root node, selecting the feature and split point that minimizes variance.
  Split the dataset into subsets based on the chosen feature and split point.
  Recursively repeat the process for each subset until a stopping criterion is met.

- Stopping Criteria:
  The tree-building process stops when a predefined maximum depth is reached.
  Alternatively, the process halts if the number of instances in a node falls below a specified threshold.

- Prediction:
  For a new instance, traverse the tree to find the leaf node to which it belongs.
  The predicted value is the mean (or median) of the target values in that leaf node.

Example:
Consider predicting house prices based on features like square footage and the number of bedrooms. The decision tree might make splits such as "If square footage <= 1500, go left; otherwise, go right."

Evaluation Metrics:

1) Mean Squared Error (MSE):
   Measures the average squared difference between predicted and actual values.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

2) Mean Absolute Error (MAE):
   Measures the average absolute difference between predicted and actual values.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

3) R-squared (Coefficient of Determination):
   Measures the proportion of the variance in the dependent variable that is predictable from the independent variables.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$