1. **Confusion Matrix**: A table used to evaluate the performance of a classification model by comparing predicted class labels to actual class labels. It provides insights into the types of errors made by the model.
2. **Error Matrix**: Another name for a confusion matrix.
3. **Contingency Table**: A table that displays the frequencies of different combinations of two categorical variables. In the case of a confusion matrix, the variables are the actual and predicted classes.
4. **True Positive (TP)**: An instance where the model correctly predicts the positive class.
5. **False Positive (FP)**: An instance where the model incorrectly predicts the positive class when the actual class is negative.
6. **False Negative (FN)**: An instance where the model incorrectly predicts the negative class when the actual class is positive.
7. **True Negative (TN)**: An instance where the model correctly predicts the negative class.
8. **Accuracy**: The proportion of correctly classified instances out of the total instances. It's calculated as (TP + TN) / (TP + FP + TN + FN).
9. **Precision**: The proportion of true positive predictions out of all positive predictions. It's calculated as TP / (TP + FP).
10. **Recall**: The proportion of true positive predictions out of all actual positive instances. It's calculated as TP / (TP + FN). Also known as sensitivity.
11. **Type I Error**: Incorrectly rejecting the null hypothesis when it is actually true. Equivalent to a false positive.
12. **Type II Error**: Incorrectly failing to reject the null hypothesis when it is actually false. Equivalent to a false negative.
13. **Underfitting**: Occurs when a model is too simple to capture the underlying patterns in the data, resulting in poor performance on both training and test data.
14. **Overfitting**: Occurs when a model is too complex and learns the training data too well, including noise. This leads to good performance on training data but poor generalization to new, unseen data.
15. **Bias-Variance Tradeoff**: The inherent tradeoff in machine learning between bias (error due to oversimplification) and variance (error due to overcomplexity). The goal is to find the right balance for optimal model generalization.
16. **Precision-Recall Tradeoff**: The tradeoff between precision and recall. Increasing precision often decreases recall, and vice versa. The optimal balance depends on the specific application and the relative costs of false positives and false negatives.
17. **F1 Score**: The harmonic mean of precision and recall. It provides a single metric that balances both precision and recall.
18. **ROC Curve**: A graphical representation of the tradeoff between true positive rate (sensitivity) and false positive rate (1-specificity) for different classification thresholds.
19. **AUC (Area Under the ROC Curve)**: A measure of the overall performance of a classification model. A higher AUC indicates better discrimination ability.