1. Задача классификации заключается в присвоении объекту одной или нескольких категорий на основе его характеристик. Например, определение, приведет ли пациент к сердечно-сосудистому заболеванию или нет, на основе его возраста, пола, артериального давления и других признаков.
2. Обучение с учителем - это процесс обучения модели на основе размеченных данных, где каждый пример обучающего набора имеет соответствующую целевую переменную (метку класса). В задаче классификации модель обучается на данных, где известны метки классов для каждого примера.
3. Разделение обучающей выборки позволяет оценить производительность модели на независимом наборе данных, который не использовался в процессе обучения. Это важно для оценки обобщающей способности модели и предотвращения переобучения.
4. Переобученная модель - это модель, которая слишком точно подстроена под обучающие данные и плохо обобщает новые данные. Это может произойти, когда модель слишком сложна или когда обучающая выборка слишком мала. Чтобы бороться с переобучением, можно использовать методы регуляризации, увеличивать размер обучающей выборки или упрощать модель.
5. Обобщающая способность модели машинного обучения - это ее способность правильно предсказывать метки классов для новых, ранее не виденных данных. Хорошая обобщающая способность означает, что модель хорошо работает на новых данных, не участвовавших в обучении.
6. Матрица ошибок - это инструмент для оценки производительности модели классификации. Она показывает количество верных и неверных классификаций для каждого класса. Матрица ошибок рассчитывается сравнением фактических меток классов с предсказанными метками классов моделью.
7. Accuracy (точность) - это доля правильно классифицированных примеров от общего числа примеров. Precision (точность) - это доля истинно положительных примеров среди всех примеров, которые модель предсказала как положительные. Recall (полнота) - это доля истинно положительных примеров, которые были правильно идентифицированы моделью из общего числа истинно положительных примеров. Эти метрики позволяют оценить различные аспекты производительности модели классификации.