1. Задача классификации в машинном обучении заключается в присвоении объектам определенных меток классов на основе их признаков. В данном случае, задача классификации заключается в предсказании риска сердечного приступа на основе доступных данных о пациентах.

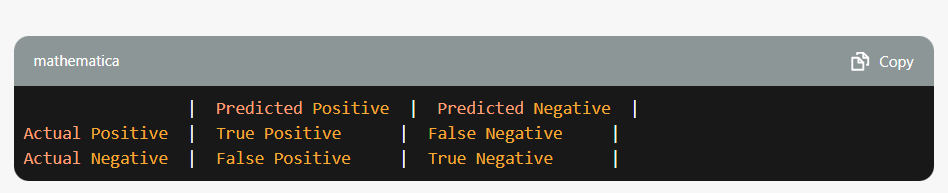
2. Обучение с учителем — это тип задачи машинного обучения, при котором модель обучается на основе помеченных данных, где каждому примеру данных соответствует известная метка класса или правильный ответ. Модель стремится выучить зависимости между входными признаками и соответствующими метками классов, чтобы в дальнейшем предсказывать метки для новых, ранее не встречавшихся данных.

3. Разделение обучающей выборки на обучающую и тестовую выборки необходимо для оценки производительности модели. Обучающая выборка используется для обучения модели, то есть определения зависимостей между признаками и метками классов. Тестовая выборка используется для оценки производительности модели на новых, ранее не встречавшихся данных. Разделение выборки позволяет оценить, насколько хорошо модель обобщает зависимости, выявленные на обучающей выборке, на новые данные.

4. Переобученная модель — это модель, которая слишком точно "запомнила" обучающие данные и не обобщает свои предсказания на новые данные. Это может привести к плохой производительности модели на тестовых данных или новых наборах данных. Для борьбы с переобучением можно использовать различные методы, такие как уменьшение сложности модели, сбор большего количества данных, регуляризация или использование методов выбора признаков.

5. Обобщающая способность моделей машинного обучения означает их способность выполнять хорошую классификацию (или регрессию) на новых, ранее не встречавшихся данных. Если модель обладает хорошей обобщающей способностью, она будет способна делать правильные предсказания на данных, которые не были использованы при ее обучении. Обобщающая способность является важным показателем качества модели и позволяет оценить ее способность к обобщению на реальные проблемы.

6. Матрица ошибок (confusion matrix) - это таблица, которая показывает количество верно и неверно классифицированных примеров для каждого класса модели. Матрица ошибок имеет следующую структуру:



True Positive (TP) - количество примеров, которые были правильно предсказаны как положительные классы.

False Negative (FN) - количество примеров, которые были неправильно предсказаны как отрицательные классы.

False Positive (FP) - количество примеров, которые были неправильно предсказаны как положительные классы.

True Negative (TN) - количество примеров, которые были правильно предсказаны как отрицательные классы.

Матрица ошибок позволят оценить производительность классификационной модели и вычислить различные метрики, такие как точность (accuracy), точность (precision) и полноту (recall).

7. Метрики accuracy (точность), precision (точность) и recall (полнота) являются мерами производительности модели классификации, вычисляемыми на основе матрицы ошибок.

Accuracy (точность) - это доля правильно классифицированных примеров от общего числа примеров. Формула для вычисления точности:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Precision (точность) - это доля правильно классифицированных положительных примеров от общего числа примеров, предсказанных как положительные. Формула для вычисления точности:

Precision = TP / (TP + FP)

Recall (полнота) - это доля правильно классифицированных положительных примеров от общего числа положительных примеров. Формула для вычисления полноты:

Recall = TP / (TP + FN)

Метрика accuracy показывает общую точность модели, метрика precision оценивает, насколько модель правильно предсказывает положительные примеры, а метрика recall оценивает, какую долю положительных примеров модель способна обнаружить. В зависимости от конкретной задачи и приоритетов можно выбирать наиболее подходящую метрику для оценки производительности модели.