

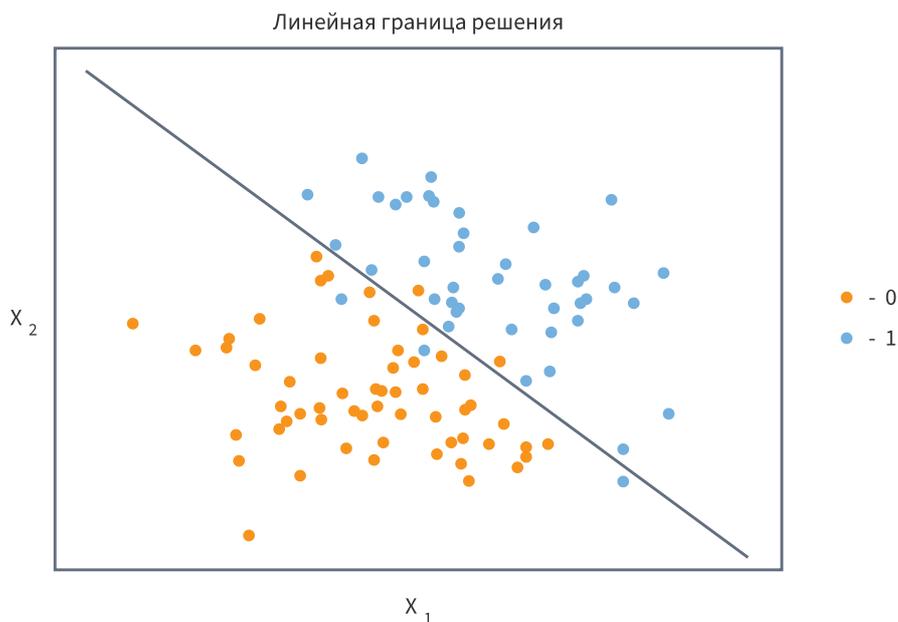
# Граница решения (Decision boundary)

Синонимы: Граница принятий решения, Решающая граница, Дискриминирующая граница

Граница решения является фундаментальным понятием в машинном обучении, а именно в классификации, и представляет собой линию или поверхность, которая разделяет пространство признаков на области, ассоциирующиеся с классами. Границы формируются в процессе обучения модели-классификатора на обучающих данных, а затем используются для классификации новых наблюдений в процессе практического использования ML-модели, которая определяет его класс в зависимости от того, в какую область он попадет.

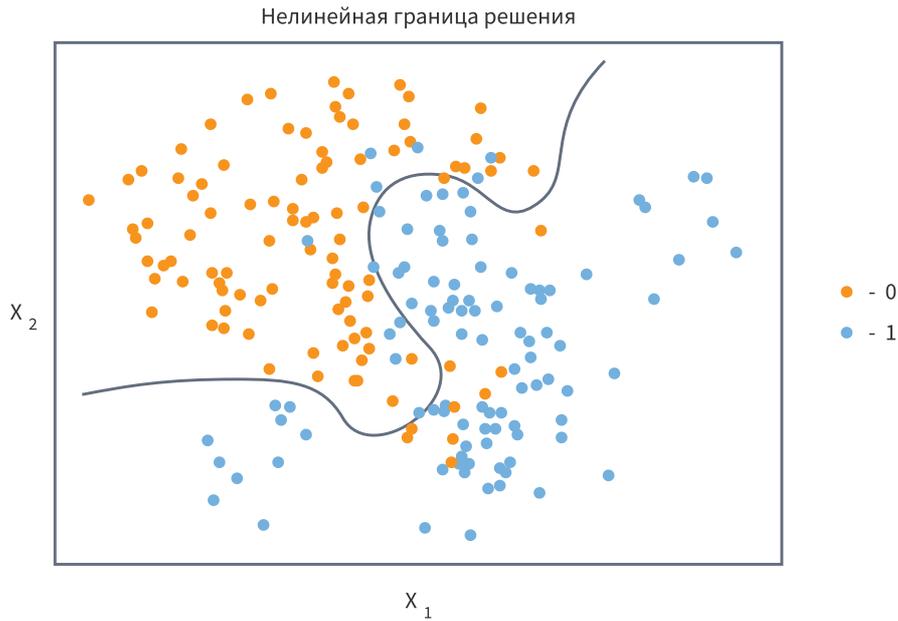
В зависимости от сложности задачи число границ и их форма различны. Для бинарной классификации используется единственная граница решения, а для мультиклассовой — несколько границ. Можно выделить следующие формы границ:

**Линейная.** Это простейшая форма границы решения, которая используется, когда классы в задаче классификации являются линейно-разделимыми. Она представляет собой линию в 2-мерном, плоскость в 3-мерном и гиперплоскость в многомерном пространствах.

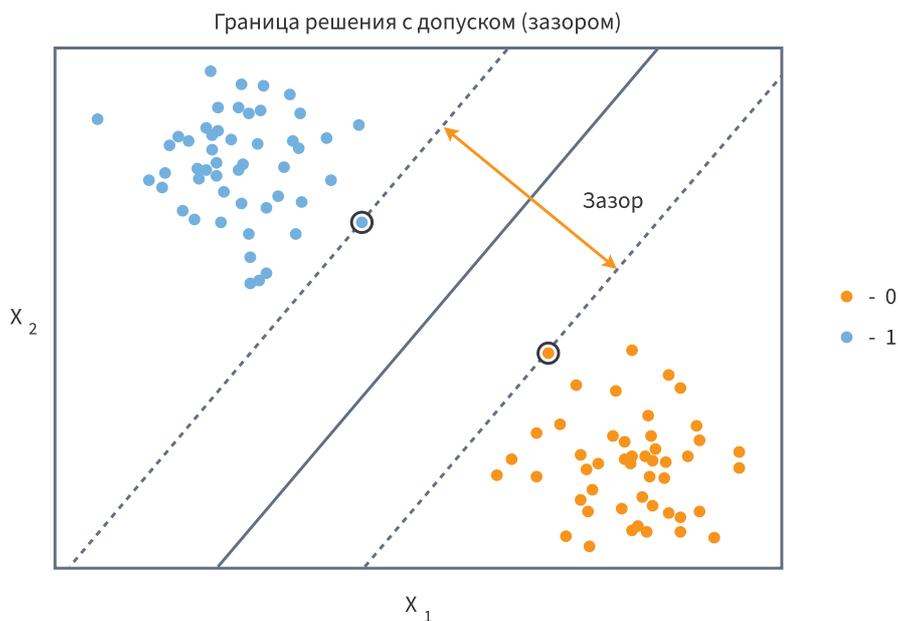


**Нелинейная.** Используется, когда классы не являются линейно-разделимыми. Нелинейные границы решений могут принимать различные формы, такие как параболы, окружности, эллипсы в 2-мерном, поверхности в 3-мерном и гиперповерхности в многомерном

пространстве.

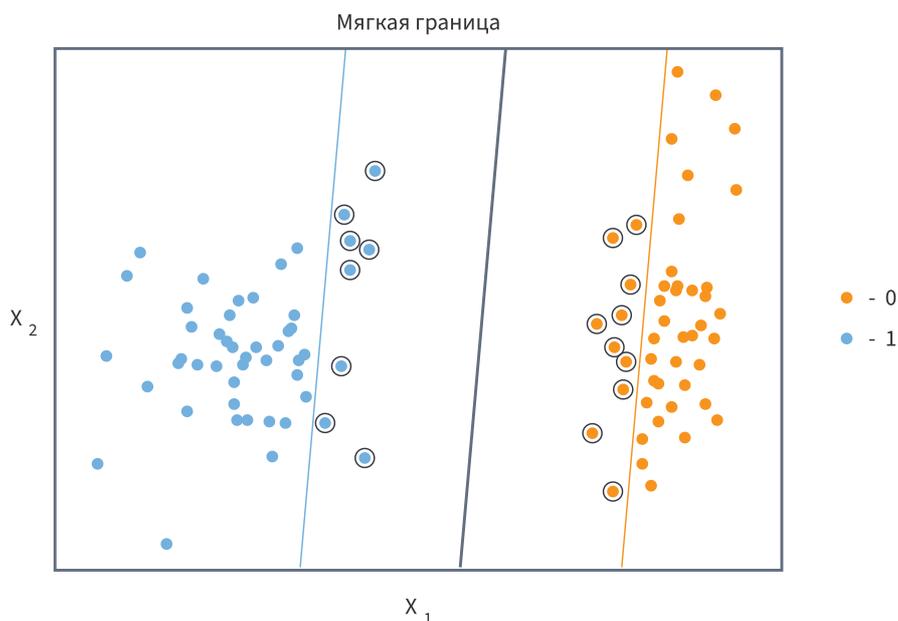


**Граница решения с допуском (зазором).** Это линия или кривая, которая разделяет данные на два класса, максимально увеличивая расстояние между границей и ближайшими точками данных. Допуск определяется как расстояние между границей решения и ближайшими точками данных каждого класса. Использование границы с допуском позволяет увеличить точность классификатора и снизить вероятность переобучения.



**Мягкая граница решения.** Это линия или кривая, которая разделяет данные на два класса, допуская при этом некоторые неверные классификации. Она используется, когда классы не являются линейно разделимыми, и когда обучающие данные для

классификатора содержат шум или выбросы. Использование мягкой границы позволяет найти баланс между точностью классификатора и его обобщающей способностью.



На рисунке представлена мягкая граница с зазором. Если жесткая граница не допускает попадания объектов внутрь зазора, то мягкая граница — допускает. Впоследствии на попавшие внутрь зазора объекты при классификации накладывается штраф, который повышает ее точность.

Границы решений формируются с помощью алгоритмов обучения классификаторов. Наиболее популярными из которых являются:

- Оценка максимального правдоподобия — статистический метод для оценки параметров распределения вероятностей на основе наблюдаемых данных. Он используется для оценки параметров модели, которая генерирует границу решения, наилучшим образом соответствующую обучающим данным.
- Машина опорных векторов — алгоритм обучения с учителем, который применяется для поиска границы решения с зазором, разделяющим данные на два или более класса. Алгоритм производит преобразование задачи классификации в пространство большей размерности, что позволяет найти лучшее решение.
- Деревья решений — это алгоритм обучения с учителем, который используется для генерации границ решения с помощью рекурсивного разделения пространства признаков на более компактные подпространства, каждое из которых ассоциируется с классом. Граница решения определяется древовидной структурой, где узлы представляют собой правила, основанные на значении признака, а листьями связаны метки класса.
- Нейронные сети — это тип алгоритма обучения с учителем, который используется для создания границ решений путем нелинейного отображения между входными векторами и выходными классами. Нейронные сети состоят из нескольких слоев нейронов, которые выполняют нелинейные преобразования входных данных для создания выходных меток классов.

С применением границ решений связаны несколько проблем, которые стоит избегать при их построении:

- Переобучение (overfitting, overtraining) имеет место, если граница слишком точно подогнана к обучающим данным, в результате чего классификатор имеет низкую обобщающую способность.
- Недообучение (underfitting) возникает, когда граница решения слишком простая и не может отразить все закономерности в данных. Недообучение можно устранить, используя более сложные алгоритмы классификации (например, нелинейные вместо линейных).
- Дисбаланс классов (class imbalance) происходит, когда обучающая выборка содержит различное распределение примеров по классам. Дисбаланс классов приводит к построению смещенных границ решения в пользу доминирующего класса. Дисбаланс можно устранить, используя корректные методы формирования выборки.

Решение перечисленных проблем обучения классификаторов является необходимым условием качественной работы модели на реальных данных.