### Формула для расчета AUC

AUC вычисляется как площадь под кривой ROC (Receiver Operating Characteristic), которая строится на основе значений истинных положительных и ложных положительных классификаций при различных порогах.

#### Шаги для расчета AUC:

1. **Построение кривой ROC:**
   * **True Positive Rate (TPR)**: Это доля правильных положительных классификаций (т.е. дефолтных заемщиков, которых модель правильно классифицировала как дефолтных).
   * **False Positive Rate (FPR)**: Это доля ложных положительных классификаций (т.е. заемщиков, которых модель ошибочно классифицировала как дефолтных).
2. **Построение ROC-кривой:** На графике, где ось X — это FPR, а ось Y — это TPR, строится кривая, которая показывает, как меняются значения TPR и FPR при различных порогах классификации.
3. **Расчет AUC:** Площадь под ROC-кривой вычисляется интегрированием функции TPR относительно FPR. Эта площадь и будет являться значением AUC:

* В реальных вычислениях AUC обычно рассчитывается с использованием численных методов, таких как метод трапеций, для аппроксимации площади под ROC-кривой.

### Простое приближенное вычисление AUC:

Для дискретных данных AUC можно вычислить по следующей формуле:

где:

* — количество положительных примеров (например, дефолтных заемщиков),
* — количество отрицательных примеров (например, заемщиков без дефолта),
* и — прогнозируемые вероятности дефолта для положительных и отрицательных примеров соответственно,
* — индикаторная функция, которая равна 1, если (т.е. если модель присвоила более высокий риск дефолта положительному примеру, чем отрицательному), и 0 в противном случае.

Это выражение подсчитывает количество всех пар из положительных и отрицательных примеров, в которых модель правильно классифицировала положительный пример с более высокой вероятностью дефолта, чем отрицательный.

### Интерпретация значения AUC:

* **AUC = 0.5**: Модель не лучше случайного классификатора, т.е. она не может отличить положительные примеры от отрицательных (кривая ROC совпадает с диагональю).
* **AUC > 0.5**: Модель имеет некоторую предсказательную способность, и чем выше значение AUC, тем лучше она различает дефолтных заемщиков от недефолтных.
* **AUC = 1**: Модель идеально классифицирует все примеры.

Для практических целей AUC обычно используется как один из ключевых индикаторов качества модели, особенно когда важен баланс между правильным и ложным распознаванием дефолтов.

AUC (Area Under the Curve) — это метрика, используемая для оценки качества классификатора, которая измеряет способность модели различать положительные и отрицательные классы. В контексте модели PD Lifetime, AUC будет оценивать, насколько хорошо модель различает дефолтных и не дефолтных заемщиков.

### Формула для расчета AUC

AUC вычисляется как площадь под кривой ROC (Receiver Operating Characteristic), которая строится на основе значений истинных положительных и ложных положительных классификаций при различных порогах.

#### Шаги для расчета AUC:

1. **Построение кривой ROC:**
   * **True Positive Rate (TPR)**: Это доля правильных положительных классификаций (т.е. дефолтных заемщиков, которых модель правильно классифицировала как дефолтных).
   * **False Positive Rate (FPR)**: Это доля ложных положительных классификаций (т.е. заемщиков, которых модель ошибочно классифицировала как дефолтных).
2. **Построение ROC-кривой:** На графике, где ось X — это FPR, а ось Y — это TPR, строится кривая, которая показывает, как меняются значения TPR и FPR при различных порогах классификации.
3. **Расчет AUC:** Площадь под ROC-кривой вычисляется интегрированием функции TPR относительно FPR. Эта площадь и будет являться значением AUC:

* В реальных вычислениях AUC обычно рассчитывается с использованием численных методов, таких как метод трапеций, для аппроксимации площади под ROC-кривой.

### Простое приближенное вычисление AUC:

Для дискретных данных AUC можно вычислить по следующей формуле:

где:

* — количество положительных примеров (например, дефолтных заемщиков),
* — количество отрицательных примеров (например, заемщиков без дефолта),
* и — прогнозируемые вероятности дефолта для положительных и отрицательных примеров соответственно,
* — индикаторная функция, которая равна 1, если (т.е. если модель присвоила более высокий риск дефолта положительному примеру, чем отрицательному), и 0 в противном случае.

Это выражение подсчитывает количество всех пар из положительных и отрицательных примеров, в которых модель правильно классифицировала положительный пример с более высокой вероятностью дефолта, чем отрицательный.

### Интерпретация значения AUC:

* **AUC = 0.5**: Модель не лучше случайного классификатора, т.е. она не может отличить положительные примеры от отрицательных (кривая ROC совпадает с диагональю).
* **AUC > 0.5**: Модель имеет некоторую предсказательную способность, и чем выше значение AUC, тем лучше она различает дефолтных заемщиков от недефолтных.
* **AUC = 1**: Модель идеально классифицирует все примеры.

Для практических целей AUC обычно используется как один из ключевых индикаторов качества модели, особенно когда важен баланс между правильным и ложным распознаванием дефолтов.