

Квантильная трансформация

Квантиль — каждый из любого набора значений вариации, которые делят распределение частот на равные группы, каждая из которых содержит одну и ту же долю от общей совокупности.

Квантильная трансформация — нелинейная трансформация данных, обеспечивающая непараметрическое преобразование, основанное на квантильной функции (2.7), для сопоставления данных с равномерным распределением со значениями от 0 до 1.

$$F^{-1}(q) \equiv F^{-}(q) = \inf \{x: F(x) \geq q\}, \quad (2.7)$$

где $q \in (0; 1)$.

Этот метод преобразует признаки так, чтобы они следовали равномерному или нормальному распределению. Следовательно, для конкретного признака это преобразование имеет тенденцию распространять наиболее частые значения, что уменьшает влияние выбросов. Таким образом, это надежная схема предобработки.

Преобразование применяется к каждому признаку независимо. Функция кумулятивной плотности (CDF) используется для прогнозирования исходных значений. Характеристические значения новых или незнакомых данных, которые опускаются ниже или выше установленного диапазона, будут отображаться в границах выходного распределения. Заметим, что это преобразование нелинейно. Оно может искажать линейные корреляции между переменными, измеренными в одном и том же масштабе, но делает переменные, измеренные в разных масштабах, более прямо сопоставимыми. Или, другими словами, подобно методам масштабирования, такая трансформация помещает каждый признак в один и тот же диапазон или распределение. Выполняя преобразование ранга, она сглаживает необычные распределения и меньше подвержена влиянию выбросов, чем методы масштабирования. Однако она искажает корреляцию и расстояние внутри и между признаками. На рисунке 2.5 показана трансформация образца импульса изолированного сигнала.