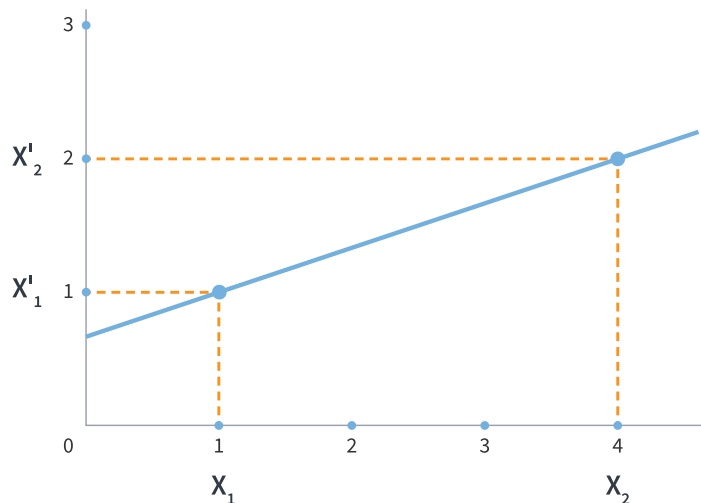


Нормирование линейное (Linear normalization)

Синонимы: Нормализация линейная

Нормирование данных заключается в приведении диапазона изменения значений признаков к некоторым требуемым границам (например, от 0 до 1). Нормирование является необходимым начальным этапом обработки данных при использовании многих многомерных статистических методов — снижения размерности признакового пространства, классификации и т.д., особенно если переменные измерены в шкалах, существенно различающихся в величинах (например, от миллиметров до километров).

Если связь между начальным и конечным диапазонами выражается линейной функцией, то нормирование является линейным (значение признаков по всему диапазону изменяются на одну и ту же величину). Если используется нелинейная функция (например, экспонента), то нелинейным (значения признаков, расположенные в различных частях диапазона, изменяются по разному).



Линейное нормирование предпочтительно в том случае, когда значения переменной X равномерно заполняют определенный интервал. Если в данных имеются редкие аномалии, намного превышающие типичный разброс, то в этих случаях следует ориентироваться при нормировке не на экстремальные (граничные) значения:

$$\tilde{x}_i = \frac{x_i - x_{i,min}}{x_{i,max} - x_{i,min}},$$

а на типичные значения — среднее:

$$\bar{x}_i = \frac{1}{P} \sum_{k=1}^P X_I^k$$

и дисперсию:

$$\tilde{x}_i = \frac{x_i - \bar{x}_i}{\sigma_i}.$$