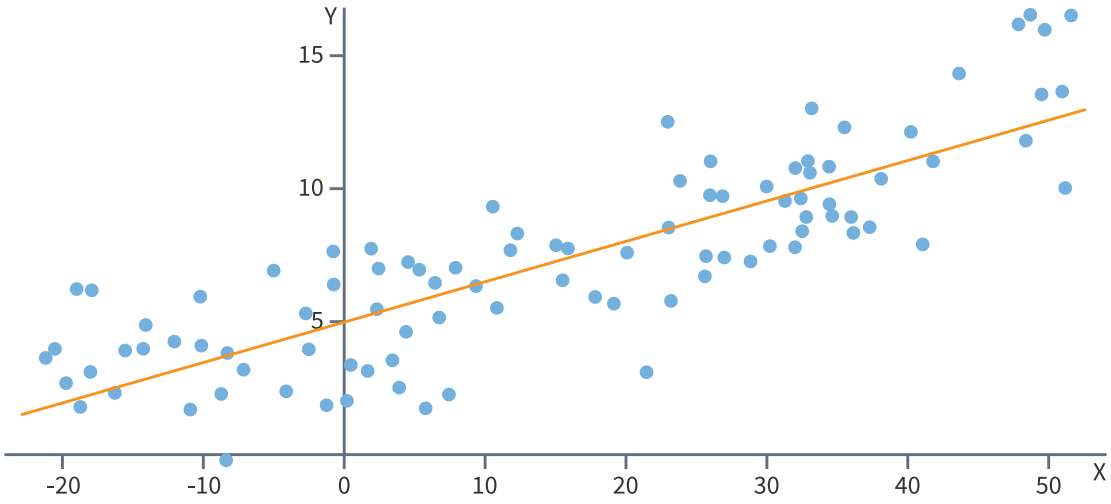
Линейная регрессия (Linear regression)

Разделы: Алгоритмы
Loginom: Линейная регрессия (обработчик)

В математической статистике линейная регрессия представляет собой метод аппроксимации зависимостей между входными и выходными переменными на основе линейной модели. Является частью более широкой статистической методики, называемой регрессионным анализом.

В регрессионном анализе входные (независимые) переменные называются также предикторными переменными или регрессорами, а зависимые переменные — критериальными.

Если рассматривается зависимость между одной входной и одной выходной переменными, то имеет место простая линейная регрессия. Для этого определяется уравнение регрессии $y = a\,x + b$ и строится соответствующая прямая, известная как линия регрессии.



Коэффициенты a и b, называемые также **параметрами модели**, определяются таким образом, чтобы сумма квадратов отклонений точек, соответствующих реальным наблюдениям данных, от линии регрессии была бы минимальной. Коэффициенты обычно оцениваются методом наименьших квадратов.

Если ищется зависимость между несколькими входными и одной выходной переменными, то имеет место множественная линейная регрессия. Соответствующее уравнение имеет вид:

$$Y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n,$$

где n — число входных переменных. Очевидно, что в данном случае модель будет описываться не прямой, а гиперплоскостью. Коэффициенты уравнения множественной линейной регрессии подбираются так, чтобы минимизировать сумму квадратов отклонения реальных точек данных от этой гиперплоскости.

Линейная регрессия была первым видом регрессионного анализа, который был тщательно изучен и начал широко использоваться в практических приложениях. Это связано с тем, что в линейных моделях оценивание параметров проще, а также с тем, что статистические свойства полученных оценок легче определить.

Линейная регрессия имеет много практических применений. Большинство приложений попадают в одну из двух широких категорий:

- Если целью является прогнозирование, линейную регрессию можно использовать для подгонки модели к наблюдаемому набору данных.
- Если цель заключается в том, чтобы объяснить изменчивость выходной переменной, можно применить линейный регрессионный анализ для количественной оценки силы взаимосвязи между выходной и входными переменными.