

УДК 004.85

## ЛИНЕЙНАЯ РЕГРЕССИЯ

Ерошевич К.В., ассистент кафедры ИиАПС  
Кузбасский государственный технический университет  
имени Т.Ф. Горбачева  
г. Кемерово

Линейная регрессия является одним из наиболее популярных алгоритмов и чаще всего используется для начала решения любой задачи регрессии. Очень часто считается простейшей моделью машинного обучения.

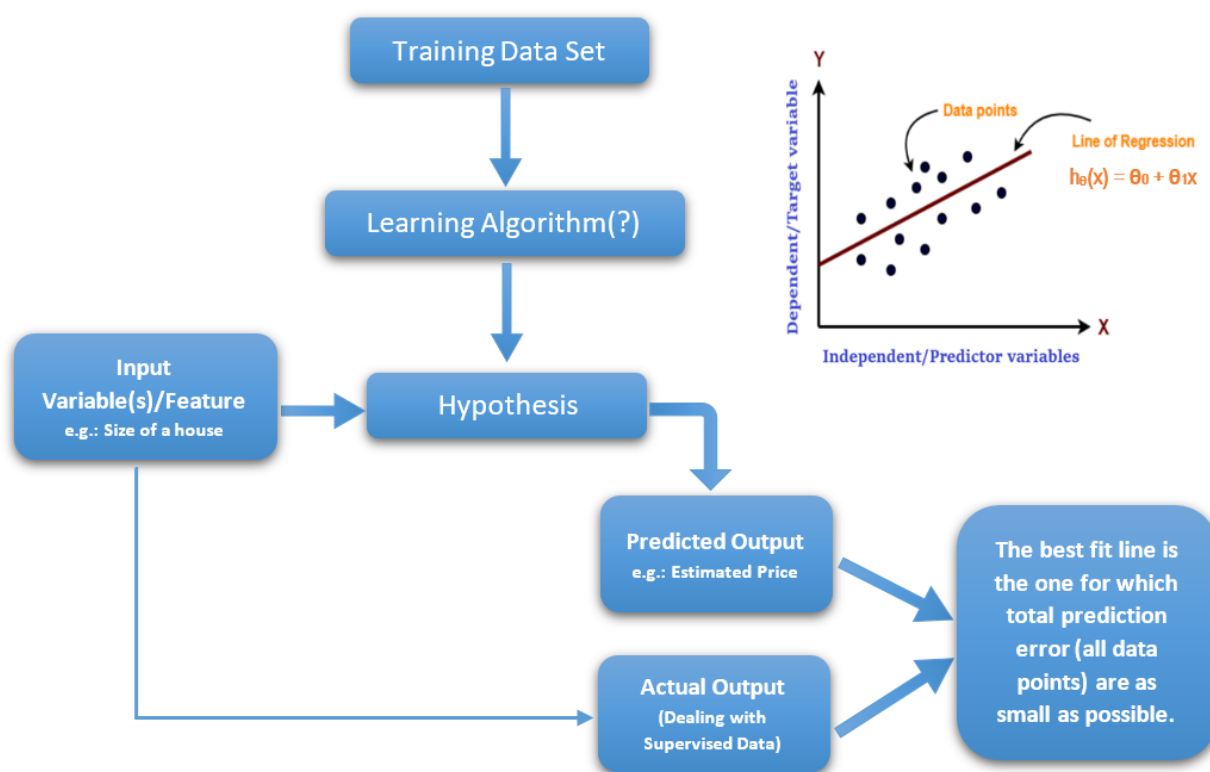


Рисунок 1. Технологическая схема модели линейной регрессии

Идея здесь заключается в том, чтобы выяснить взаимосвязь между зависимой / целевой переменной ( $y$ ) для одной или нескольких независимых переменных ( $x$ ) в наборе обучающих данных. Применяя алгоритм обучения, который в нашем случае представляет собой линейную регрессию, использующую линейную функцию в качестве гипотезы. Чтобы решить

параметры этого уравнения гипотезы, нужно минимизировать функцию затрат (также известную как функция квадрата ошибки для задач регрессии). Эти минимизированные параметры находятся с помощью метода градиентного спуска. Окончательное уравнение, полученное на предыдущем шаге, используется для прогнозирования будущего.

Некоторые из его приложений могут прогнозировать продажи определенного товара на основе его цены, сезона или других параметров. Он также может быть использован в анализе временных рядов для прогнозирования результатов обработки данных для прогнозирования будущего.

## Теория

Само слово *Линейная регрессия* имеет три производных термина: Линейный, Линейный и Регрессионный.

Линия - это кратчайшее расстояние между двумя точками. В то время как линейный относится к набору точек, лежащих на прямой линии на плоскости, отображающей непрерывные значения.

Регрессия - это форма метода прогностического моделирования, который исследует взаимосвязь или корреляцию между зависимой и независимой переменной.

Что такое независимая переменная?

Это именно то, на что это похоже. Это переменная, часто называемая целевой переменной, которая стоит отдельно и не изменяется другими переменными. Например, чей-то возраст может быть независимой переменной. Другие факторы не изменят возраст человека.

На самом деле, когда вы ищете какую-то взаимосвязь между переменными, вы пытаетесь увидеть, вызывает ли независимая переменная какие-либо изменения в других переменных или зависимых переменных.

Что такое зависимая переменная?

Зависимая переменная - это переменная, которая зависит от других факторов. Например, результат теста может быть зависимой переменной, потому что он может изменяться в зависимости от нескольких факторов, таких как то, как много вы учились, сколько вы спали ночью перед сдачей теста или даже насколько вы были голодны, когда сдавали его. Обычно, когда

ищется взаимосвязь между двумя вещами, то выясняется, что заставляет зависимую переменную изменяться так, как она меняется.

Знание уравнения (уравнений)

Поскольку мы можем определить целевую переменную и иметь метки для них, эта проблема относится к области контролируемой области, где данные уже помечены.

The diagram shows the equation  $y = mx + c$  with color-coded labels and arrows pointing to each part:

- $y$  is labeled "dependent variable" (blue arrow).
- $m$  is labeled "Slope" (red arrow).
- $x$  is labeled "independent variable" (brown arrow).
- $c$  is labeled "y intercept" (blue arrow).

Рисунок 2. Уравнение прямой

Итак, здесь взаимосвязь линейной регрессии лучше всего определяется уравнением прямой линии, которое также является гипотезой линейной регрессии и также известно большинству из нас из математики средней школы, и задается :

$$y = m * x + c$$

В машинном обучении часто используется такой вид уравнения:

$$h(x) = w_0 + w_1 * x_1$$

или

$$h_0(x) = \theta_1 * x_1 + \theta_0$$

где

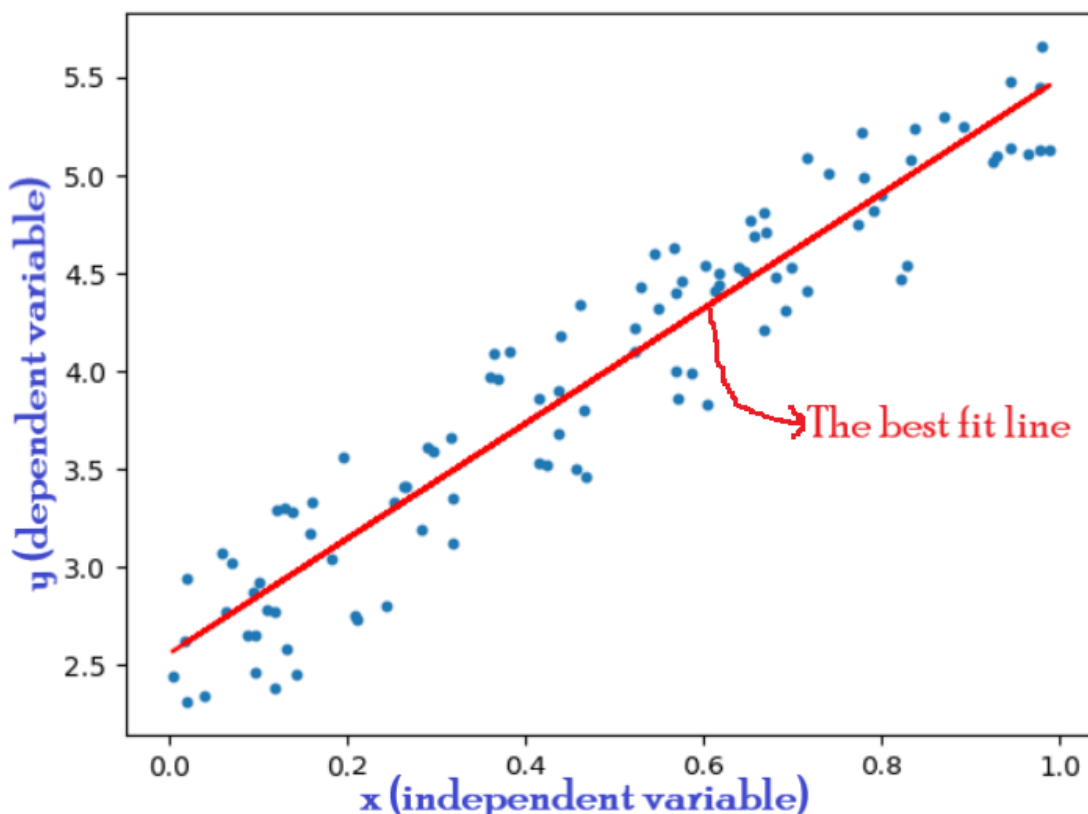
1.  $y$  или  $h(x)$  - зависимая или целевая переменная
2.  $m$  или  $w_1$  или  $x_1$  - уклон или наклон
3.  $x$  или  $x_1$  - зависимая или предикторная переменная
4.  $c$  или  $w_0$  или  $\theta_0$  - пересечение по оси  $y$

Это уравнение поможет связать зависимую и независимую переменную, также известную как простая линейная регрессия, а как насчет нескольких зависимых переменных? Его можно определить с помощью множественной линейной регрессии и обобщить следующим образом:

$$h_0(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots + \theta_n x_n$$

Что такое Линейная регрессия?

Простая линейная регрессия - это статистический метод, который позволяет нам обобщать и изучать взаимосвязь между двумя реальными переменными. Простыми словами, найдите строку для описания линейной функции, которая наилучшим образом представляет данные, а также некоторое время оценивается как «Линия наилучшего соответствия»(the best fit line).



**Рисунок 3. Пример наилучшей подходящей линии**

Основываясь на заданных точках данных, мы пытаемся найти строку, которая наилучшим образом описывает данные. Красная линия на

приведенном выше графике выполняет эту работу за нас и также называется линией наилучшего соответствия

Давайте определим приведенную выше линию наилучшего соответствия математически с помощью уравнения:

$$h_0(x) = \theta_0 + \theta_1 * x$$

где  $\theta_0$  и  $\theta_1$  - действительные числа, которые описывают линейное уравнение наилучшего соответствия. Как найти эти реальные цифры и когда можно сказать, что это лучшее? Мы начинаем с некоторого предположения об этих действительных числах, а затем пытаемся получить разницу в погрешности между средними прогнозируемыми и фактическими точками данных. Это делается путем минимизации функции затрат, а затем предполагаемые действительные числа обновляются с использованием метода градиентного спуска, пока не достигнем оптимального решения.

### Список литературы

1. Машинное обучение – это легко [Электронный ресурс] // Хабр: Сообщество IT-специалистов – Режим доступа: <https://habr.com/ru/post/319288/> – (Дата обращения: 01.02.2022).
2. Введение в линейную регрессию для машинного обучения [Электронный ресурс] // КодКамп – Режим доступа: <https://codecamp.ru/blog/introduction-to-linear-regression-python/> – (Дата обращения: 11.02.2022).
3. Линейная регрессия в машинном обучении [Электронный ресурс] // neurohive – Режим доступа: <https://neurohive.io/ru/osnovy-data-science/linejnaja-regressija/> – (Дата обращения: 10.02.2022).