

Лекция 3. Нелинейная регрессия

Введение

До сих пор мы предполагали, что мир устроен линейно: "увеличение X на единицу всегда ведет к постоянному изменению Y ". Но так ли это? Представьте:

- Увеличение дохода на 1000 рублей для студента и для миллионера по-разному влияет на их потребление.
- Эффект от дополнительного часа учебы для экзамена зависит от того, учились ли вы до этого 2 часа или 10.

Сегодня мы узнаем, как моделировать такие **нелинейные по своей природе** зависимости, часто используя хорошо знакомый нам аппарат **линейной регрессии**.

1. Зачем нужны нелинейные модели?

1.1. Экономическая теория

Многие экономические законы по своей сути нелинейны. Классический пример — **закон убывающей предельной отдачи**. Первый работник на ферме увеличивает урожай сильно, пятый — еще немного, а пятидесятый может даже мешать. Эффект от дополнительной единицы фактора производства (X) зависит от его текущего уровня.

1.2. Более адекватное описание данных

Данные часто имеют явно нелинейный паттерн. Попытка провести через них прямую линию может привести к грубым ошибкам в предсказаниях и неверным выводам.

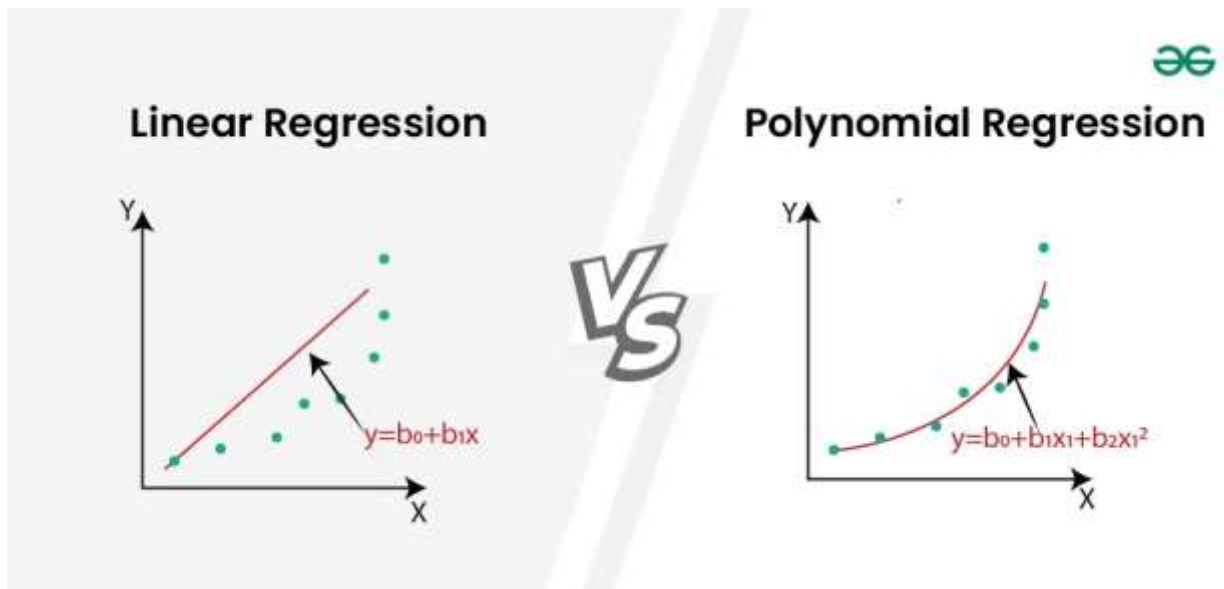


Рисунок 3.1. Линейная и нелинейная регрессия

На графике слева видно (Рисунок 3.1), как линейная модель (прямая красная линия) плохо соответствует данным, в то время как нелинейная кривая (красная) на графике справа идеально их описывает.

1.3. Повышение точности прогноза

Учет нелинейности позволяет делать более точные прогнозы, особенно при экстраполяции за пределы диапазона данных.

2. Ключевая идея: «линеаризуемые» модели

Самая мощная идея в этой теме - многие нелинейные зависимости можно свести к линейным с помощью подходящего **преобразования переменных**.

Мы будем рассматривать модели, **нелинейные по переменным, но линейные по параметрам**. Это значит, что параметры (β_0 , β_1 , ...) по-прежнему входят в модель аддитивно и линейно. Это позволяет нам использовать все тот же надежный метод наименьших квадратов (МНК)!

3. Основные типы линеаризуемых нелинейных моделей

3.1. Полиномиальная регрессия

Используется, когда зависимость имеет изгиб (например, рост, а затем спад).

- **Модель (полином 2-й степени):** $Y_i = \beta^0 + \beta^1 X_i + \beta^2 X_i^2 + \varepsilon_i$
- **Как линеаризовать?** Мы создаем новые переменные!
 - Пусть $Z_1 = X$
 - Пусть $Z_2 = X^2$
 - Теперь модель: $Y_i = \beta^0 + \beta^1 Z_1 + \beta^2 Z_2 + \varepsilon_i$ - это обычная **множественная линейная регрессия!**
- **Интерпретация:** Интерпретация коэффициентов в полиномиальной модели усложняется. Эффект X на Y теперь зависит от значения самого X . Его называют **предельным эффектом** и вычисляют как производную:
 - $\partial Y / \partial X = \beta_1 + 2\beta_2 X$
 - *Пример:* если Y - урожай, X - количество удобрений, и $\beta_2 < 0$, то предельная отдача удобрений убывает.

3.2. Модели с логарифмированием («Лог-Модели»)

Это, пожалуй, самый популярный класс моделей в эконометрике. Они позволяют моделировать постоянные *относительные* изменения. Основные виды моделей с логарифмированием приведены в таблице 3.1.

Таблица 3.1

Название Модели	Формула	Линеаризация	Интерпретация β_1
Лог-Лин	$\ln(Y) = \beta_0 + \beta_1 X + \varepsilon$	Уже линейна	При увеличении X на 1 ед., Y изменяется в среднем на $(\beta_1 * 100)\%$
Лин-Лог	$Y = \beta_0 + \beta_1 \ln(X) + \varepsilon$	Уже линейна	При увеличении X на 1%, Y изменяется в среднем на $(\beta_1 / 100)$ ед.
Лог-Лог	$\ln(Y) = \beta_0 + \beta_1 \ln(X) + \varepsilon$	Уже линейна	При увеличении X на 1%, Y изменяется в среднем на $\beta_1\%$ (Эластичность!)

Что такое эластичность? Это процентное изменение одной переменной в ответ на процентное изменение другой. В модели "Лог-Лог" коэффициент β_1 непосредственно интерпретируется как эластичность Y по X. Это невероятно удобно для экономистов.

- *Пример (Лог-Лог):* Модель спроса: $\ln(Q) = \beta_0 + \beta_1 \ln(P) + \varepsilon$, где Q — количество, P — цена. Коэффициент β_1 — это ценовая эластичность спроса.

4. Общий алгоритм работы и предостережения

Шаги для построения модели:

1. **Гипотеза:** исходя из теории или визуального анализа данных (постройте диаграмму рассеяния!) выдвигайте гипотезу о форме зависимости.
2. **Преобразование:** создайте новые переменные (X^2 , $\ln(X)$, $\ln(Y)$ и т.д.).

3. **Оценка:** оцените параметры преобразованной модели с помощью МНК.
4. **Интерпретация:** Интерпретируйте коэффициенты в соответствии с типом модели.

Важные замечания:

- **Предсказания в исходной шкале:** если вы преобразовывали Y (например, моделировали $\ln(Y)$), то предсказанное значение \hat{Y} получается не простым потенцированием, а с поправкой на смещение, если мы хотим предсказывать именно Y , а не $\ln(Y)$. Это важно!
- **R^2 несравнимы:** Коэффициент детерминации R^2 для моделей с разным зависимой переменной (например, Y и $\ln(Y)$) **нельзя сравнивать напрямую**. Они измеряют долю объясненной дисперсии в разных переменных.
- **Гомоскедастичность:** Преобразования (особенно логарифмирование) часто помогают стабилизировать дисперсию ошибок, если изначально она росла с ростом X .

5. Пример: Зарплата и опыт (Полиномиальная модель)

Теория: Зарплата растет с опытом, но с определенного момента рост замедляется (предельная отдача от опыта убывает).

Модель:

$$\text{Зарплата}_i = \beta^0 + \beta^1 * \text{Опыт}_i + \beta^2 * \text{Опыт}_i^2 + \varepsilon_i$$

Оценка модели (условные числа):

$$\hat{Y} = 30 + 5 * \text{Опыт} - 0.1 * \text{Опыт}^2$$

Интерпретация:

- Предельный эффект опыта: $\partial Y / \partial X = 5 - 0.2 * \text{Опыт}$.
- Для человека с 10 годами опыта: $5 - 0.2 * 10 = 3$. Значит, дополнительный год опыта увеличивает зарплату на 3 тыс. у.е.

- Для человека с 20 годами опыта: $5 - 0.2 \cdot 20 = 1$. Рост всего на 1 тыс. у.е.
- Мы даже можем найти "оптимальный" опыт (вершину параболы), после которого зарплата начинает падать.

Резюме

1. **Не все зависимости линейны.** Использование нелинейных моделей продиктовано экономической теорией и видом данных.
2. **Многие нелинейные модели линеаризуемы.** Путем преобразования переменных (квадраты, логарифмы) мы можем свести задачу к множественной линейной регрессии.
3. **Интерпретация коэффициентов меняется кардинально.** В нелинейных моделях мы говорим о **предельных эффектах** и **эластичностях**.
4. **Модель "Лог-Лог"** дает прямую оценку **эластичности** - одному из ключевых понятий в экономике.

На следующей лекции: Мы сделаем следующий логический шаг и перейдем к **множественной линейной регрессии**, где мы будем изучать совместное влияние нескольких факторов на зависимую переменную, что является основным инструментом для выявления причинно-следственных связей.

Вопросы для самопроверки:

1. Представьте, что вы изучаете зависимость цены квартиры от ее площади. Какую модель вы бы предпочли: Лин-Лог или Лог-Лог? Почему?
2. Оценили модель: $\ln(\text{Цена}) = 8 + 1.5 \cdot \ln(\text{Площадь})$. Как интерпретировать коэффициент 1.5?
3. В полиномиальной модели $Y = \beta_0 + \beta_1 X + \beta_2 X^2$ коэффициент β_2 оказался положительным. Какой будет форма кривой? О чем это может говорить?

