

ЛЕКЦИИ 3. Полиномиальная регрессия: анализ нелинейных зависимостей

Введение

В первых двух лекциях вы научились строить простую регрессионную модель и узнали, как метод наименьших квадратов подбирает наилучшую прямую линию, минимизируя ошибки.

Такая модель полезна, когда одна переменная линейно влияет на другую: например, больше экранного времени → хуже сон, или выше доход → больше удовлетворённость.

Но сегодня мы задаём более глубокий вопрос:

всегда ли реальный мир ведёт себя линейно?

Ответ — **нет**.

Почему “прямая линия” не всегда работает

В реальной жизни зависимости часто **непрямые**.

До определённого момента — переменная действительно помогает.

Но потом её влияние может замедляться, останавливаться, а то и становиться **обратным**.

Примеры:

Подготовка к экзамену:

До 6 часов — успеваемость растёт.

После 10 часов — утомление, тревожность, результат падает.

Доход и счастье:

До определённого уровня дохода — вырастает чувство стабильности.

После — эффект плато: деньги уже не приносят того же уровня удовлетворения.

Физическая активность:

Без неё — вялость.

Слишком много — усталость и перетренированность.

Где-то между — оптимум.

Как это связано с гипотезой?

Когда вы формулируете гипотезу, вы интуитивно предполагаете не только **есть ли связь**, но и **какая она по форме**:

Прямая?

Сначала вверх, потом вниз?

Плато?

Порог?

Если вы подозреваете **оптимальную точку** или **нелинейный поворот**, то **прямая линия регрессии не сможет это отразить**.

К чему это ведёт?

Это приводит нас к необходимости новой модели — **полиномиальной регрессии**, где в уравнение добавляется вторая степень переменной (x^2).

Это маленькое изменение — **добавить x^2** — даёт вам возможность отразить реальность гораздо точнее, чем простая линия.

Что будет дальше в лекции

Сразу после этого введения мы:

1. **Разберём формулу полиномиальной регрессии**
2. **Посмотрим реальные кейсы с "горкой" и "плато"**
3. **Научимся строить модель с x^2 в Excel и Google Sheets**
4. **Интерпретируем коэффициенты**, включая "изгиб"
5. Разберёмся, когда **можно**, а когда **нельзя** использовать полиномиальную модель

Теперь мы переходим к основной части лекции.

Скопировать расширенное введение можно полностью — оно идеально вписывается в курс и задаёт рамку для изучения **нелинейных зависимостей через полиномиальную регрессию**.

Формула полиномиальной регрессии

Чтобы учесть “изгиб” в зависимости, в модель добавляется переменная x^2 (x в квадрате):

$$y = b_0 + b_1 * x + b_2 * x^2$$

Где:

x — независимая переменная,

x^2 — квадрат этой переменной,

b_1 — линейный эффект (прямая часть зависимости),

b_2 — “изгиб” (показывает наличие пика, провала и т.д.),

b_0 — базовое значение y , когда $x = 0$.

Пример гипотезы

Гипотеза:

Существует оптимальное количество часов подготовки к экзамену, при котором студент показывает наилучший результат.

Слишком мало — знаний не хватает.

Слишком много — выгорание и снижение эффективности.

Пример данных

Студент	Часы подготовки (x)	x^2	Итоговая оценка (y)
1	1	1	55
2	3	9	70
3	5	25	85
4	7	49	88
5	10	100	82
6	12	144	75
7	15	225	60

Вы можете добавить колонку x^2 в Google Sheets/Excel формулой $=A2^2$.

Как применить модель

Google Sheets (английский интерфейс):

```
=LINEST(D2:D8, A2:B8, TRUE, TRUE)
```

Excel (русский интерфейс):

```
=ЛИНЕЙН(D2:D8; A2:B8; ИСТИНА; ИСТИНА)
```

Где:

A — x (часы подготовки),

B — x^2 ,

D — y (итоговая оценка).

Пример результата

Допустим, вы получили:

$$b_2 = -0.3$$

$$b_1 = 5.2$$

$$b_0 = 50$$

$$R^2 = 0.91$$

Модель:

$$\text{Оценка} = 50 + 5.2 * x - 0.3 * x^2$$

Интерпретация:

$b_1 = 5.2$: каждое дополнительное время подготовки сначала повышает оценку,

$b_2 = -0.3$: но с ростом часов — эффект замедляется и затем уходит в минус,

$b_0 = 50$: базовая оценка без подготовки,

$R^2 = 0.91$: 91% изменений в оценке объясняются моделью → модель сильная.

Другие кейсы для полиномиальной модели

📌 1. Доход и удовлетворённость

Доход до 500 000 ₹ — уровень удовлетворённости растёт

После — эффект плато (коэффициент $b_2 \approx 0$ или отрицательный)

📌 2. Физическая нагрузка и энергия

От 0 до 5 тренировок в неделю — энергия растёт

После 7 тренировок — появляется усталость → энергия падает

Когда использовать полиномиальную регрессию?

Ситуация	Модель подходит?
Данные образуют "горку" или "провал"	✅ Да
Зависимость "вверх → вниз"	✅ Да
Связь строго линейная	❌ Нет
Вы не знаете, как кодировать	❌ Нет

Построение графика

1. Выделите x и y
2. Постройте **точечную диаграмму**
3. Добавьте **трендовую линию**
4. Выберите тип: **Полиномиальная**, степень = 2
5. Включите отображение уравнения и R^2

Использование ИИ

Инструмент	Зачем использовать
Excel Copilot	Автоанализ тренда, построение кривой
Google Sheets Explore	Построит график и предложит модель
Notion AI	Поможет описать поведение переменных
Data Analyst GPT	Объяснит, как работает x^2 и как это читать

🚫 Запрещено:

Добавлять x^2 , если вы не видите "поворота" в данных

Не интерпретировать b_2 — это главный показатель "изгиба"

Подбирать “нужный результат” вручную

Использовать модель без визуальной проверки графика

Вывод

Полиномиальная регрессия позволяет выйти за рамки “прямых связей” и начать анализировать **реальные паттерны поведения**: где данные растут, где замедляются, где переворачиваются.

Это важный этап в развитии аналитика: вы начинаете видеть **форму зависимости**, а не просто цифры.