Машинное обучение

Лекция 2. Линейные модели. Градиентный спуск

(10.02.2024)

Даниил Литвинов

Общие сведения

План

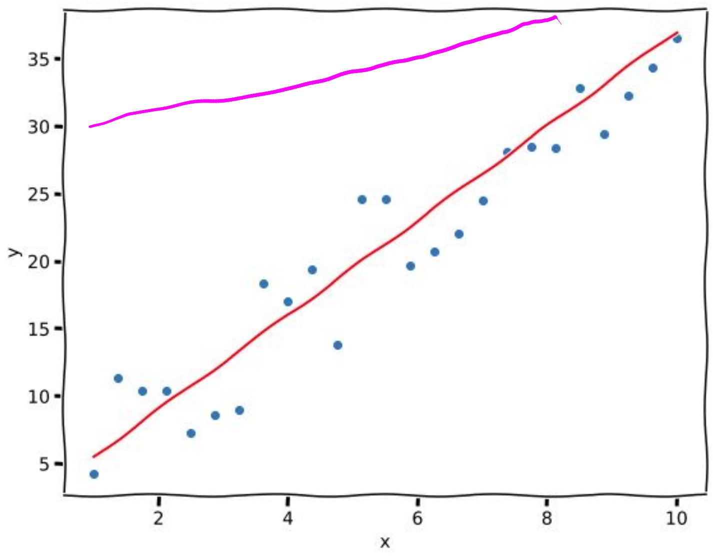
1. Линейная модель регрессии
2. Как линейные модели обучаются?
3. Линейная модель классификации

Что это такое?

x — баллы за экзамен по английскому 1

y — баллы за экзамен по английскому 2

x	y
1	5
3	11
9	35
10	33



Что это такое?

$\hat{y} = w_0 + w_1 x$

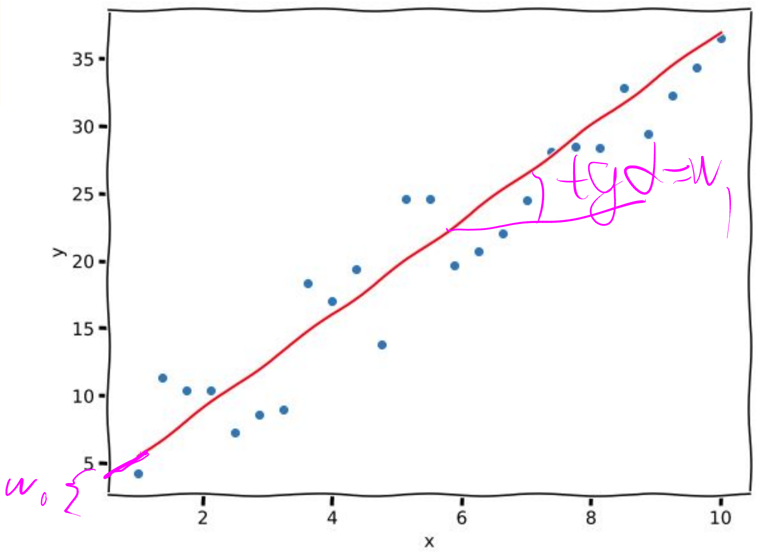
свободный член
(intercept)

случайная ошибка
(error)

$$y = w_0 + w_1 x + \epsilon$$

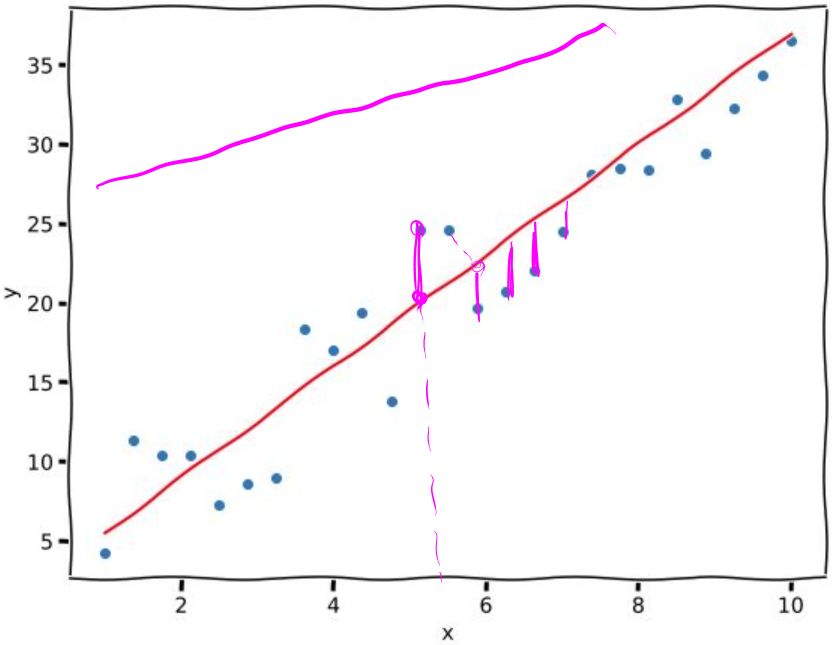
целевая переменная
(aka target)

независимая переменная
(slope, predictor)



А какая модель нам нужна?

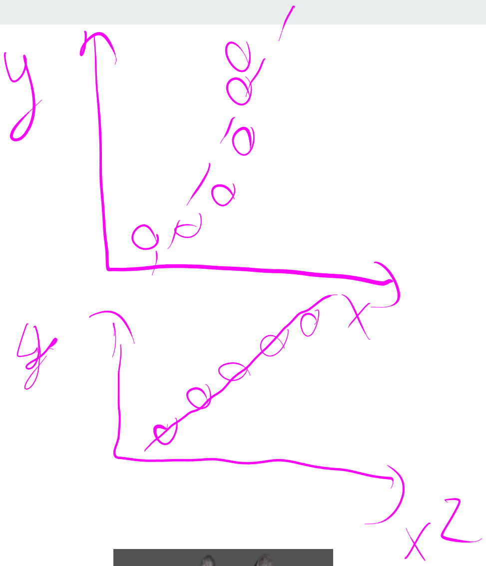
$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \to \min$$



Интерпретация коэффициентов

Зачем нужны линейные модели?

1. Предсказание интересующей нас величины
2. Оценка влияния различных факторов на нашу целевую переменную
3. Линейные модели очень легко использовать и интерпретировать
4. Линейные модели могут восстанавливать даже **нелинейные зависимости**

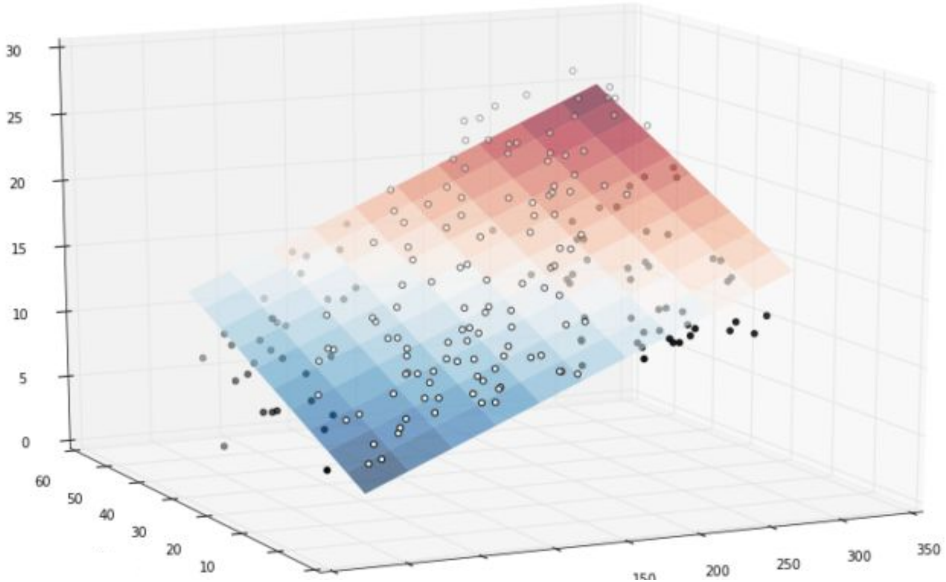


А если у нас много независимых переменных?

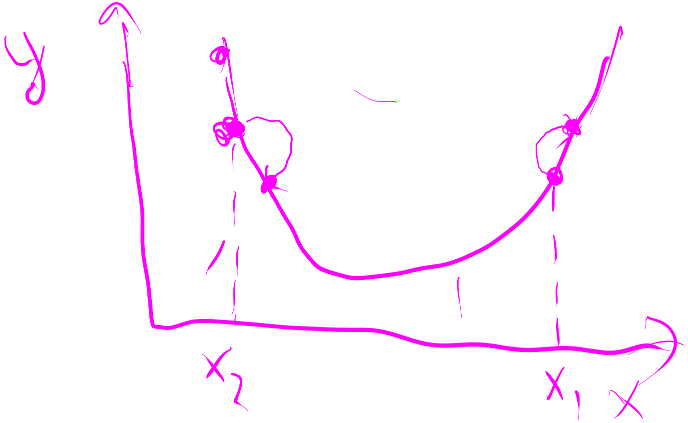
$$y = w_0 + w_1 x_1 + w_2 \overset{x_2}{\cancel{z}} + \ldots + w_n \overset{x_n}{\cancel{t}} + \epsilon$$

площадь	число комнат	школа близко	цена квартиры
50	2	нет 0	5000
1000	7	да 1	11000
30	1	нет	3500
100	4	нет	33333

Множественная линейная регрессия дает нам плоскость



Производные



$$y \sim x$$

$$\frac{dy}{dx} = \frac{y(x_0 + \Delta x) - y(x)}{\Delta x}$$

$y = f(x)$	$\frac{dy}{dx} = f'(x)$
k, any constant	0
x	1
x^2	$2x$
x^3	$3x^2$
x^n, any constant n	nx^{n-1}
e^x	e^x
e^{kx}	ke^{kx}
$\ln x = \log_e x$	$\frac{1}{x}$
$\sin x$	$\cos x$
$\sin kx$	$k \cos kx$
$\cos x$	$-\sin x$
$\cos kx$	$-k \sin kx$

Производные

$$\varphi(x,y,z) = 2x^2 + 3y^2 - \sin z$$

$$\frac{\partial \varphi}{\partial x} = 4x$$

$$\frac{\partial \varphi}{\partial y} = 6y$$

$$\frac{\partial \varphi}{\partial z} = -\cos z$$

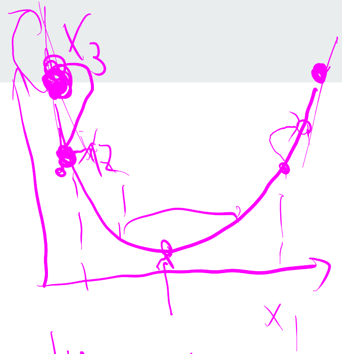
$$\begin{pmatrix} 4x \\ 6y \\ -\cos z \end{pmatrix}$$

градиент

$\nabla \varphi$ – градиент

Производные

- Град. указ. в напр. наиб. роста ф-ции.
- Антиград. ~ наоборот.



$\frac{dy}{dx}(x_1) > 0$

$\frac{dy}{dx}(x_2) < 0$

$\cancel{x_3 + \frac{dy}{dx}(x_3)}$

$x_3 - \frac{dy}{dx}(x_3)$

Производные

Как оценивать коэффициенты модели?

$$\hat{y} = w_0 + w_1 x_1 + \ldots + w_3 x_3$$

$$\underset{n\times 3}{X},\quad \underset{n\times 1}{y},\quad \underset{4\times 1}{w}$$

$$\begin{array}{cccc} 1 & x_1 & x_2 & x_3 \end{array}$$

$$\underset{n\times 1}{y} \approx \overset{n\times 3}{X} \cdot \overset{4\times 1}{w}$$

$$\rightarrow \underset{n\times 1}{y} = \overset{n\times 4}{X} \cdot \overset{4\times 1}{w}$$

$$\rightarrow \begin{bmatrix} 1 & x_1 & x_2 & x_3 \\ & & & \\ & & & \end{bmatrix} \cdot \begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{pmatrix} =$$

$$\underset{n\times 1}{\begin{bmatrix} 1\cdot w_0 + x_1 \cdot w_1 + \ldots + x_3 \cdot w_3 \\ \\ \end{bmatrix}} = \begin{pmatrix} y_1 \\ \\ \end{pmatrix} y$$

Как оценивать коэффициенты модели?

$$\hat{y} = X \cdot w$$

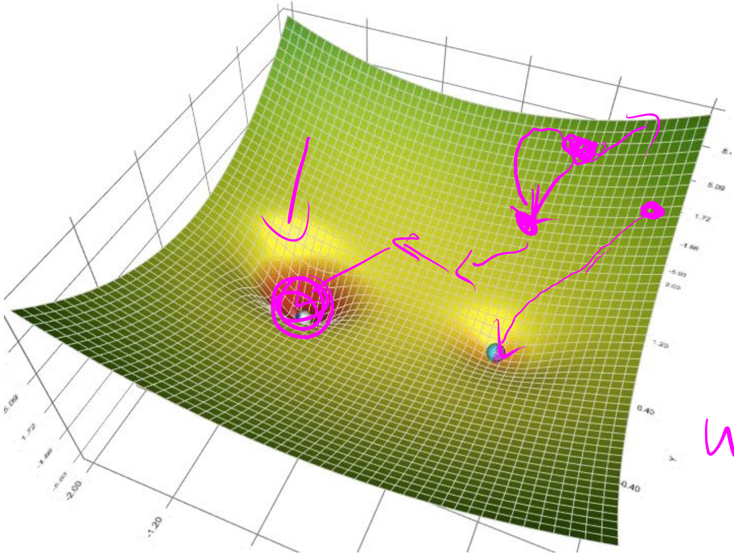
$$MSE = \frac{1}{N} (\underset{n\times 1}{y} - \underset{n\times 4}{X} \underset{4\times 1}{w})^T \cdot (y - Xw) \to \boxed{\text{функция потерь} \quad \text{Loss}}$$

$$\cancel{(n \times 1 \cdot 1 \times n \to n \times n)}$$

$$\underset{4\times 1}{DMSE_w} = -\frac{1}{N} \cdot 2 \cdot \underset{n\times 4}{X^T} \cdot (\underset{n\times 1}{y} - Xw) = \boxed{\frac{2}{N} X^T (Xw - y)}$$

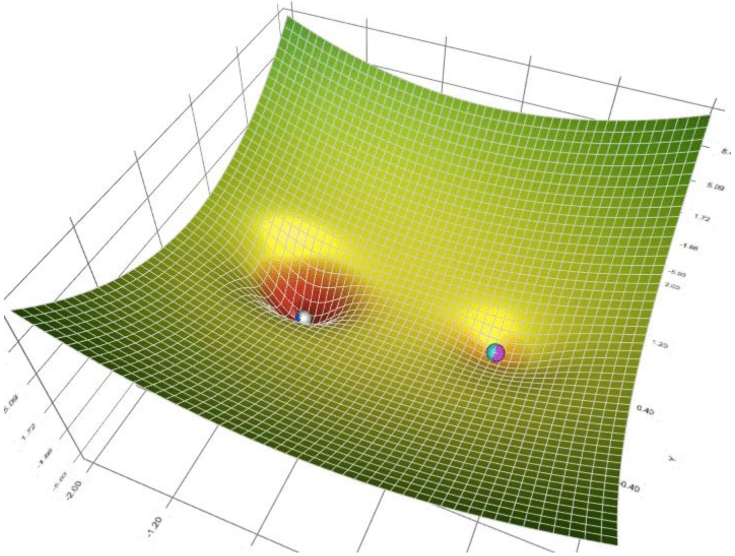
Градиентный спуск

MSE



w_1

w_0



Формулы

$$y = w_0 + w_1 x + \epsilon$$

$$y = Xw$$

$$\frac{dLoss}{dw} = \nabla Loss = 2X^T(Xw - y)$$

$$Loss = (y - Xw)^T(y - Xw)$$

$$\boxed{w = (X^T X)^{-1} X^T y}$$

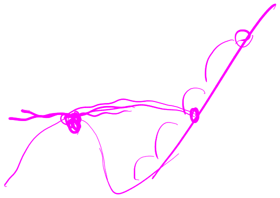
Градиентный спуск

$$Loss = (y - Xw)^T(y - Xw) \qquad \frac{dLoss}{dw} = \nabla Loss = 2X^T(Xw - y)$$

```
w = np.random.randn(m + 1)
Пока grad(Loss) != 0:
        w -= η * grad(Loss)
```

η – learning rate

[0,001, 0,01]



Отдых -> логистическая регрессия

Связь событий и признаков

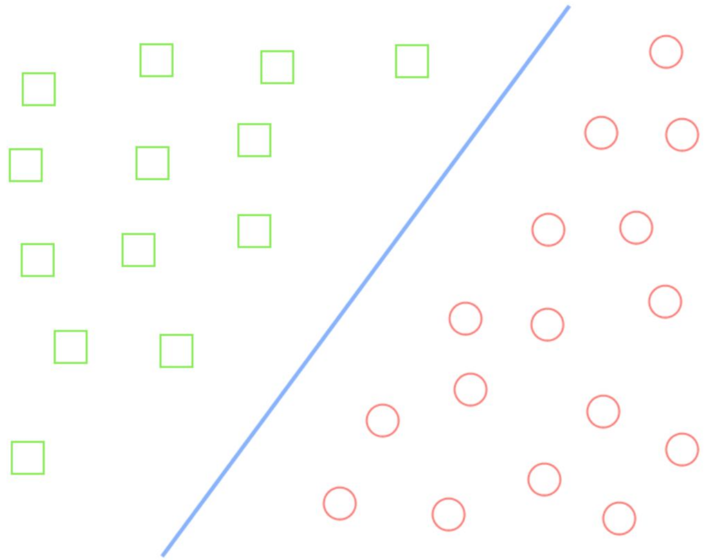
В зависимости от предикторов события могут происходить чаще или реже – логика, совпадающая с логикой связи количественной переменной отклика с набором предикторов.

Например, по мере роста температуры воздуха летом чаще будут встречаться люди в шортах: событие “встретился человек в шортах” положительно связано с температурой воздуха.

Событие “проведение исследования” явно связана с предиктором “объем полученного финансирования”, однако эта связь может быть совсем непростой.

А что если хотим классификацию?

Допустим бинарная классификация



[геометрия ML]

Отношение шансов

Шансы (odds) часто представляют в виде отношения шансов (odds ratio)

Если отношение шансов > 1, то вероятность наступления события выше, чем вероятность того, что оно не произойдет.

Если отношение шансов < 1, то наоборот.

Если можно оценить вероятность положительного события, то отношение шансов выглядит так:

$$odds = \frac{\pi}{1-\pi}$$

Отношение шансов варьируется от 0 до +∞.

Попробуем сами

Логиты

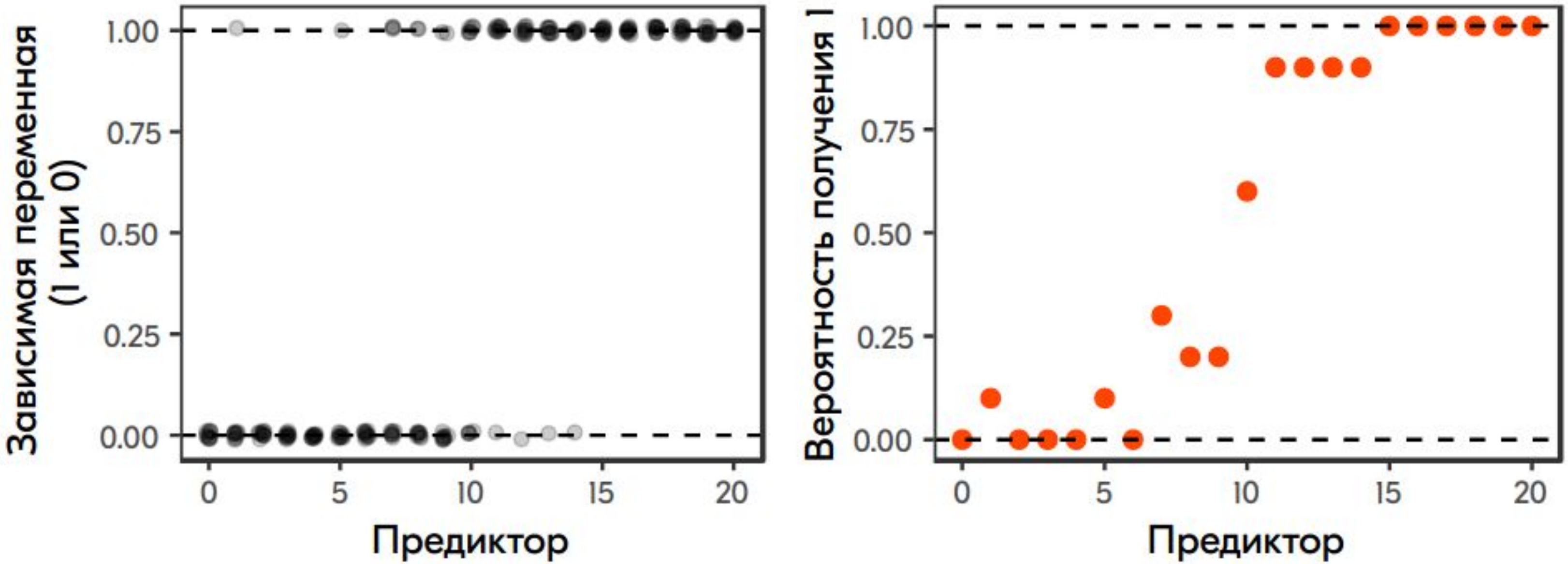
Отношение шансов можно преобразовать в логиты(logit):

$$ln(odds) = ln(\frac{\pi}{1-\pi})$$

- Значения логитов – это трансформированные оценки вероятности события.
- Логиты варьируют от $-\infty$ до $+\infty$.
- Логиты симметричны относительно 0, т.е. $ln(1)$.
- Для построения моделей в качестве зависимой переменной удобнее брать логиты.

Считаем вероятность

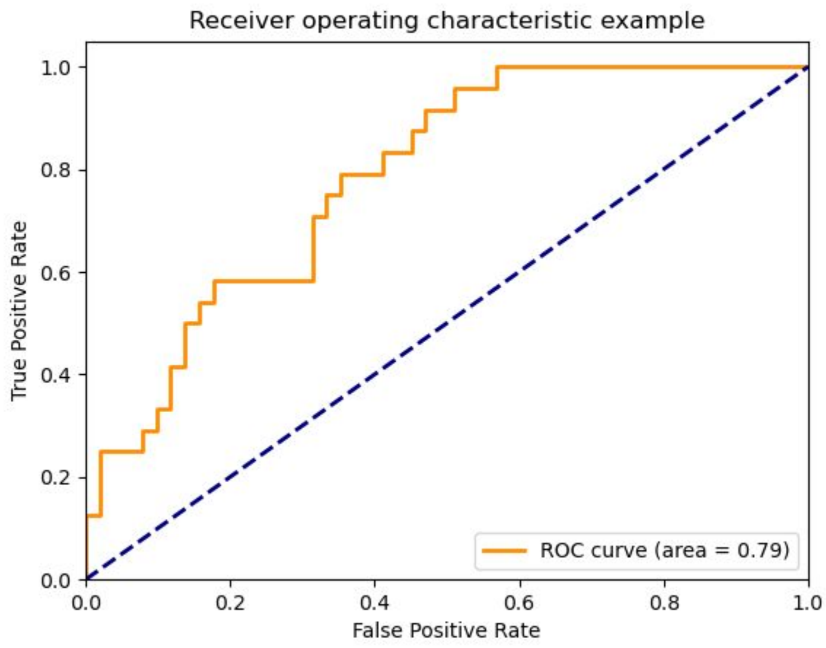
Дискретные значения *vs* вероятности



Как такое учить? BCE Loss

Качество классификации

Качество классификации. ROC кривая



[рисуем свою ROC кривую]

Построение ROC кривой