Машинное обучение

Лекция 2. Линейные модели. Градиентный спуск

(11.02.2023)

Общие сведения

План

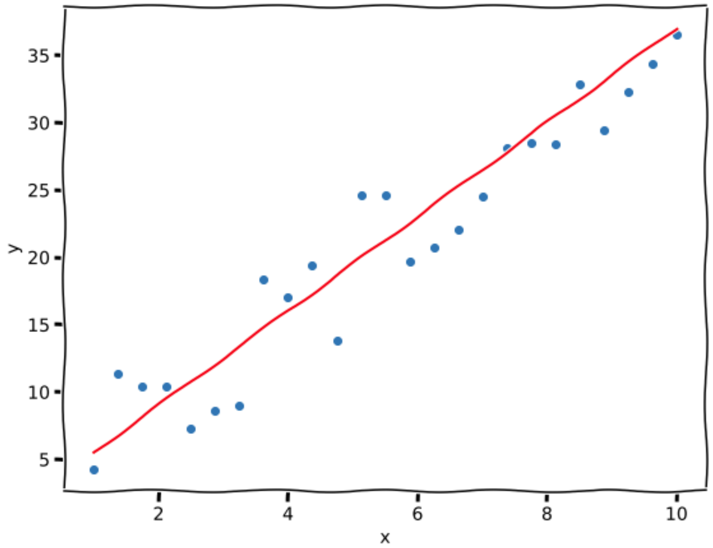
1. Линейная модель регрессии
2. Как линейные модели обучаются?
3. Линейная модель классификации

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?

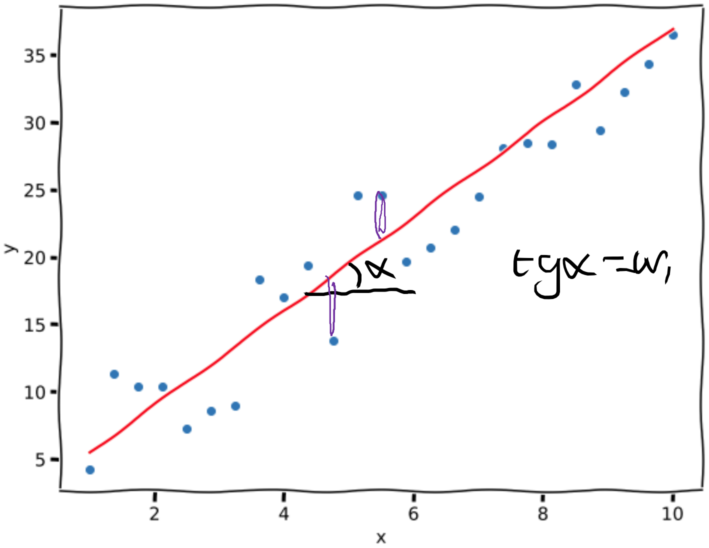
свободный член
(intercept)

случайная ошибка
(error)

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

целевая переменная
(aka target)

независимая переменная
(slope, predictor)



А какая модель нам нужна?

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

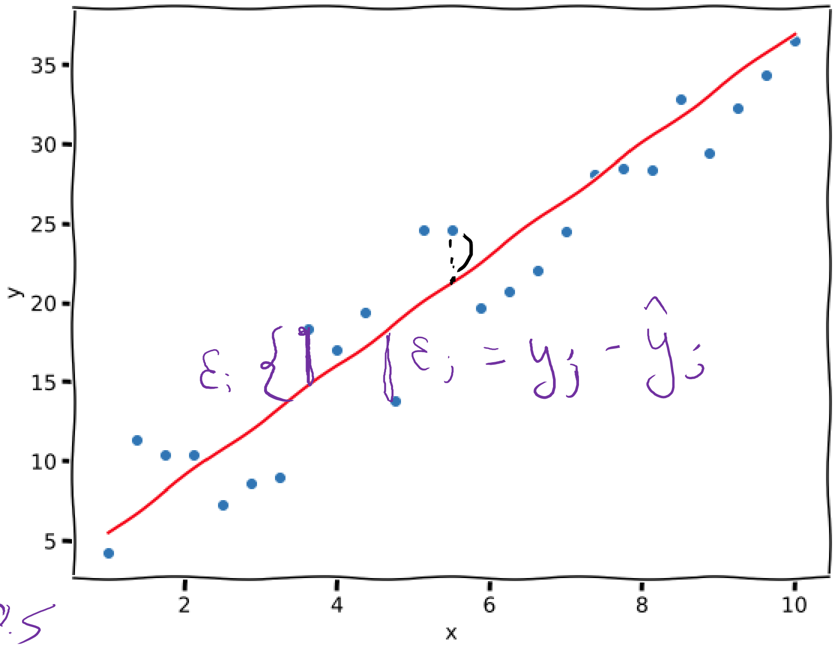
$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \to \min$$

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

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

$$y - \hat{y} = \varepsilon \to \min$$

$$\text{custom\_MSE} = \begin{cases} \varepsilon > 0: \ \varepsilon^2 \cdot 0.5 \\ \varepsilon < 0: \ \varepsilon^2 \cdot 10 \end{cases}$$



~~Интерпретация коэффициентов~~ Стандартизация

$$x_i := \frac{x_i - \bar{x}}{\sigma(x)} \to N(0,1)$$

1) kNN ?

2) $y = w_0 + w_1 \cdot \text{рост} + w_2 \cdot \text{возраст}$

y - вес, кг

$\underset{\text{кг}}{\text{кг}} = w_0 + \underset{\frac{\text{кг}}{\text{см}}}{w_1} \cdot \text{см} + \underset{\frac{\text{кг}}{\text{лет}}}{w_2} \text{лет}$

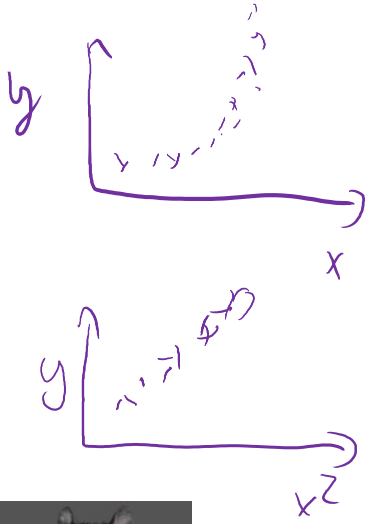
$y \sim 10^3$

$x_1 \sim 10^5$; $w_1 \sim 10^{-2}$

$x_2 \sim 10^1$; $w_2 \sim 10^2$

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**

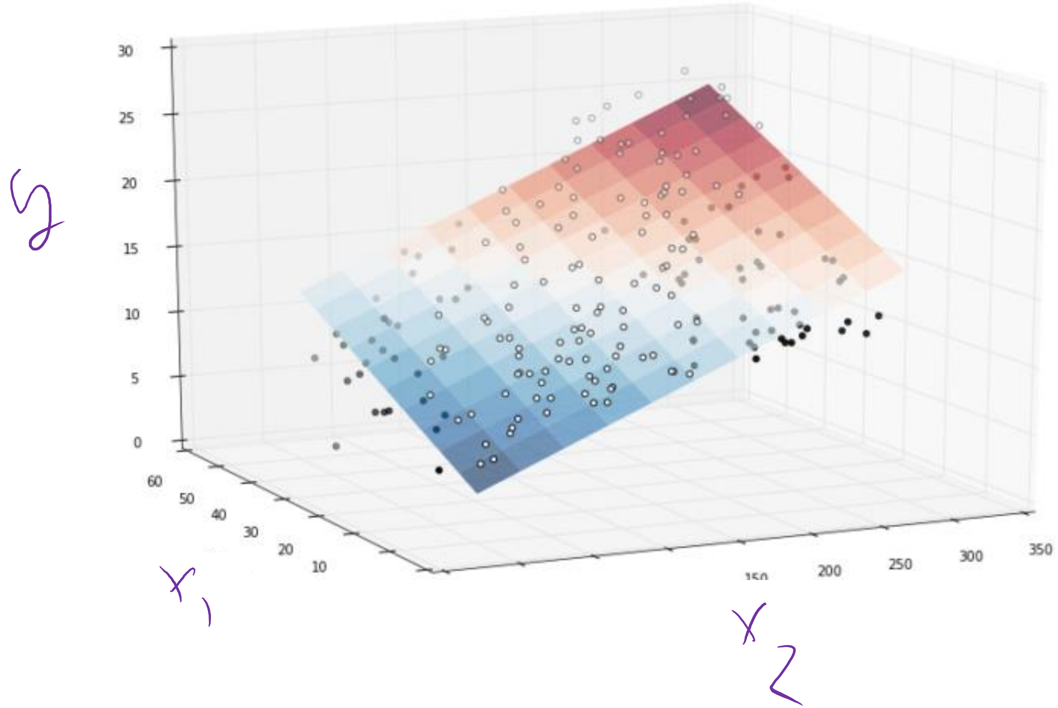


А если у нас много независимых переменных?

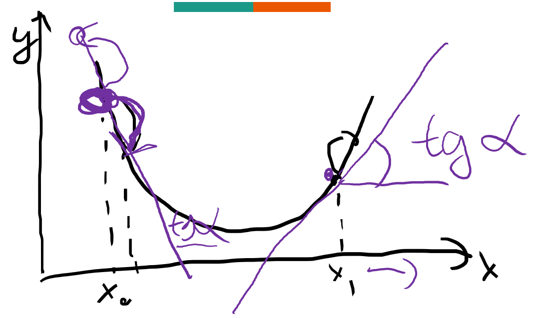
$$y = w_0 + w_1 x + w_2 z + \ldots + w_n t + \epsilon$$

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

Множественная линейная регрессия дает нам плоскость



Производные



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

1) $x_1 + \frac{dy}{dx}(x_1)$

$\downarrow$

y - увел.

2) $x_0 + \frac{dy}{dx}(x_0)$

$\downarrow$

y - увел

$\Downarrow$

$x_1 - \frac{dy}{dx}(x_1) \Rightarrow$ y - уменьш

$$\text{min/max} = \frac{x_i - \text{min}}{x_{max} - x_{min}}$$

$y = f(x)$	$\frac{dy}{dx} = f'(x)$
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

$(0, 0, \frac{\pi}{2})$ - ответ

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

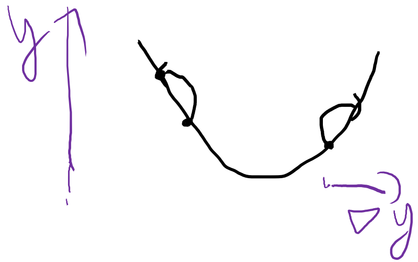
$$\frac{\partial \varphi}{\partial x} = 4x$$

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

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

$$\begin{pmatrix} 4x \\ 6y \\ -\cos z \end{pmatrix}$$ — градиент $\varphi = \nabla_{\underline{xyz}} \varphi$

Производные (градиент)



- указывает в напр. наиб. роста ф-ции ✓
- Антиградиент указ. в напр. ~~наиб.~~ наискор. убыв.

$$\begin{pmatrix} x \\ b \\ z \end{pmatrix} - \begin{pmatrix} \frac{\partial y}{\partial x} \\ \frac{\partial y}{\partial b} \\ \frac{\partial y}{\partial z} \end{pmatrix} = \begin{pmatrix} x - \frac{\partial y}{\partial x} \\ b - \frac{\partial y}{\partial b} \\ z - \frac{\partial y}{\partial z} \end{pmatrix}$$

$$y_0 > y_1$$

Производные

Как оценивать коэффициенты модели?

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

2) $\underline{y = w_0 \cdot 1 + w_1 x_1 \ldots w_3 x_3}$

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

$n \times 4$

$\underset{n\times 4}{X}$

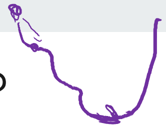
bias trick

4) $$\boxed{\underset{n\times 1}{y} = \underset{n\times 4}{X} \cdot \underset{4\times 1}{w}}$$

5) $$\underset{X}{\begin{bmatrix} 1 & x_1 & x_2 & x_3 \\ & & & \\ & & & \end{bmatrix}} \underset{w}{\begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{pmatrix}} = \underset{y}{\begin{pmatrix} \cdot \\ \\ \\ \end{pmatrix}}$$

$$= 1 \cdot w_0 + x_1 w_1 + \ldots + x_3 w_3 = y$$

Как оценивать коэффициенты модели?



функция потерь

Loss-функция → min

1) $$MSE = \frac{1}{N}(\underset{n\times1}{y} - \overset{\text{matrix}}{X}\overset{\text{веса}}{\underset{n\times1}{w}})^2 = \frac{1}{N}(\underset{n\times1}{y - Xw})^T(\underset{n\times1}{y - Xw})$$

2) $$\underset{4\times1}{\nabla_w MSE} = \boxed{-\frac{1}{N}\cdot 2\cdot \underset{4\times n}{X^T}(\underset{n\times1}{y - Xw})} = \boxed{\frac{2}{N}\cdot X^T(Xw - y)}$$

$4\times n \quad n\times4\cdot4\times1 - n\times1$

$4\times1 \sim w$

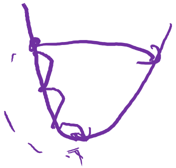
$(\quad)\cdot(\quad) \to 1$

$1\times n \quad n\times1$

$$\underset{4\times1}{w_{(i+1)}} = \underset{4\times1}{w_{(i)}} - \eta\,\nabla MSE(w_i)$$

eta - learning rate

η: [0.001, 0.01]

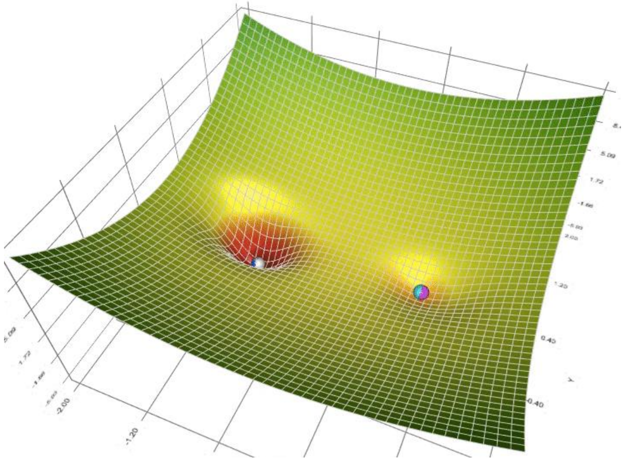


$(\,)\times(\,) \to n\times n$

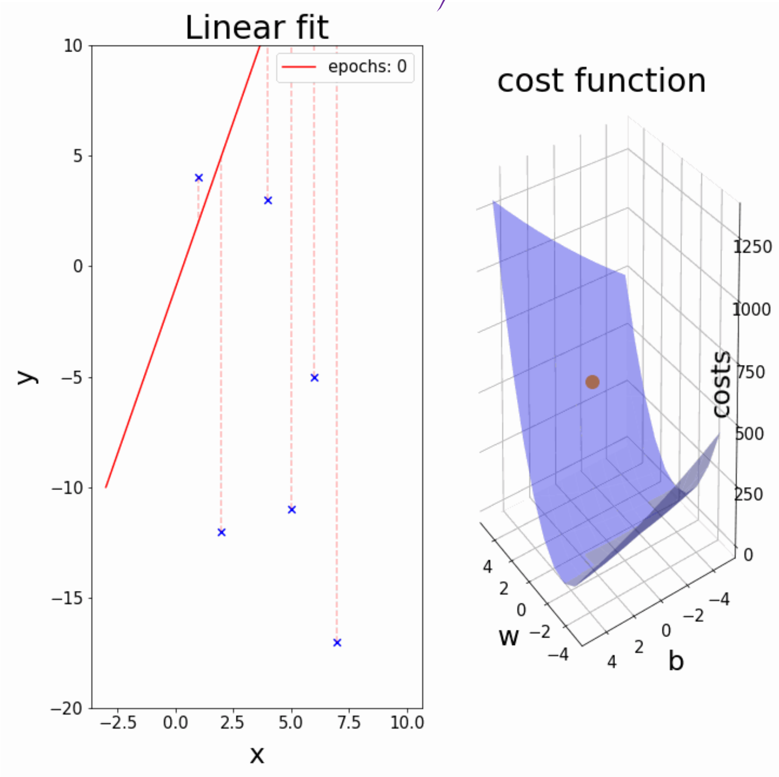
$\boxed{X @ w}$

Градиентный спуск

1)



2)



Формулы

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

$$y = Xw$$

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

OLS, МНК

$$Loss = \frac{1}{N}(y - Xw)^T(y - Xw)$$

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

Градиентный спуск

$$Loss = \frac{1}{N}(y - Xw)^T(y - Xw) \qquad \frac{dLoss}{dw} = \nabla Loss = \frac{2}{N}X^T(Xw - y)$$

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

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.

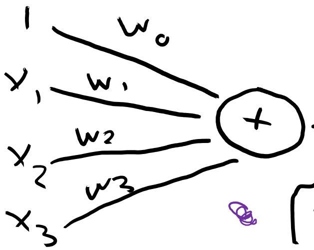
Событие “проведение исследования” явно связана с предиктором “объем полученного финансирования”, однако эта связь может быть совсем непростой.

А что если хотим классификацию?

$y = \{0, 1\}$

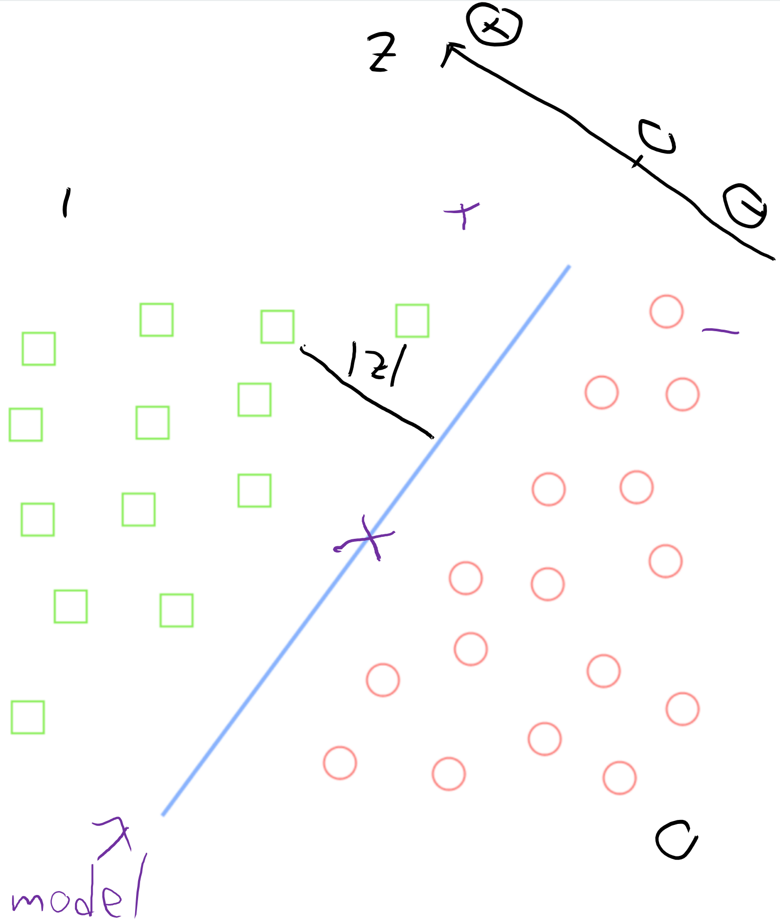
Допустим бинарная классификация

1)



$\to z \quad (-\infty; +\infty)$

- [если z больше] ⊕, то $y = 1$
- [ ] ⊖, то $y = 0$



[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:

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

Отношение шансов варьируется от 0 до +∞.

Попробуем сами

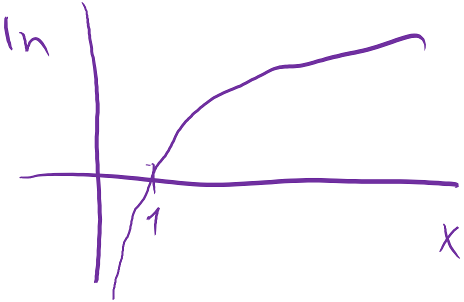
z $(-\infty; +\infty)$

1) p_i [вер. того, что] $y_i = 1$

$1 - p_i$ [ ] $y_i = 0$

2) $\dfrac{p_i}{1-p_i}$: $[0; +\infty)$

3) $\log\left(\dfrac{p_i}{1-p_i}\right)$: $(-\infty; +\infty)$



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

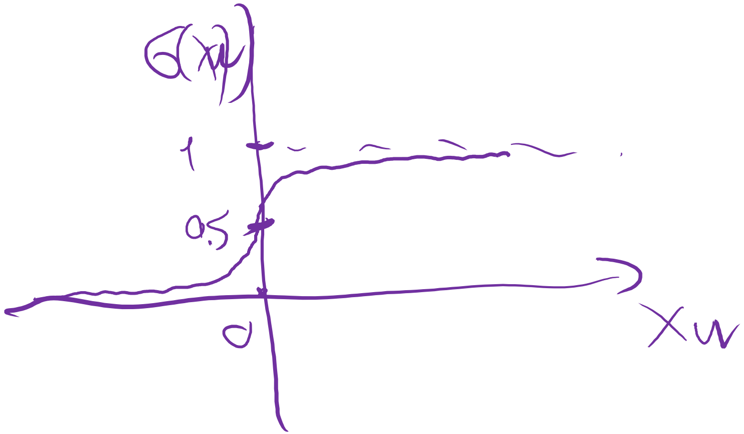
$$\ln\left(\frac{p_i}{1-p_i}\right) = Xw \quad \Big| \exp$$

$$\frac{p_i}{1-p_i} = e^{Xw}$$

$$p_i = e^{Xw} - p_i e^{Xw}$$

$$p_i(1 + e^{Xw}) = e^{Xw}$$

$$p_i = \frac{e^{Xw}}{1+e^{Xw}} = \frac{1}{e^{-Xw}+1} = \sigma(Xw)$$



Как такое учить? BCE Loss

$$BCE = \frac{1}{N}\sum_{i=1}^{N} -\left[y_i \cdot \ln p_i + (1-y_i)\cdot \ln(1-p_i)\right]$$

$y_i = \{0, 1\}$

$p_i \in [0, 1]$ $\quad p_i = \sigma(x_i\, w)$

1) $y_i = 1;\ p_i = 1;\ BCE = 0$

2) $y_i = 1;\ p_i = 0;\ BCE = +\infty$

3) $y_i = 1;\ p_i = 0{,}5;\ BCE = 0.7$

Как такое учить? BCE Loss

$$\nabla_w BCE = \boxed{X^T \cdot (p - y)}$$

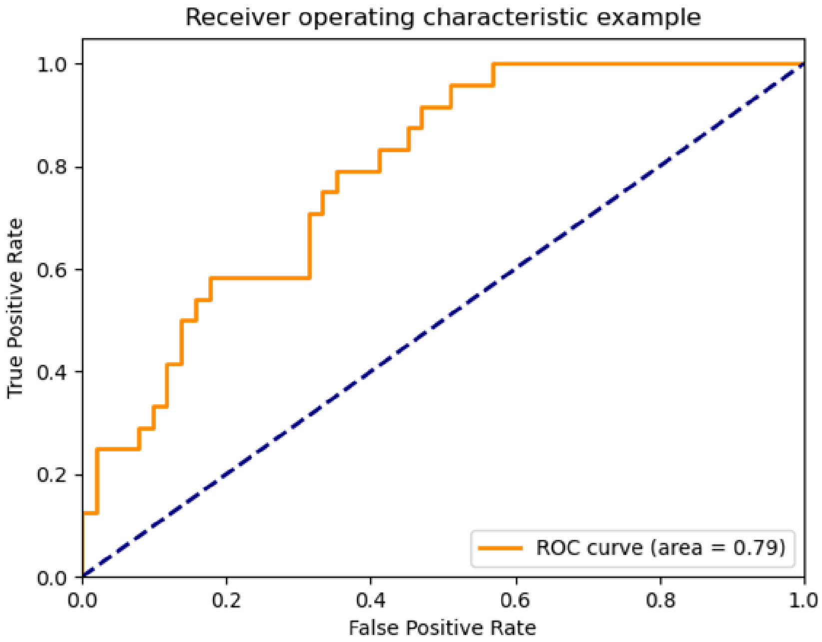
$$p = \sigma(Xw)$$

Качество классификации

Качество классификации. ROC кривая

$$TPR = \frac{TP}{TP+FN} = recall$$

$$FPR = \frac{FP}{FP+TN}$$



[рисуем свою ROC кривую]

Построение ROC кривой