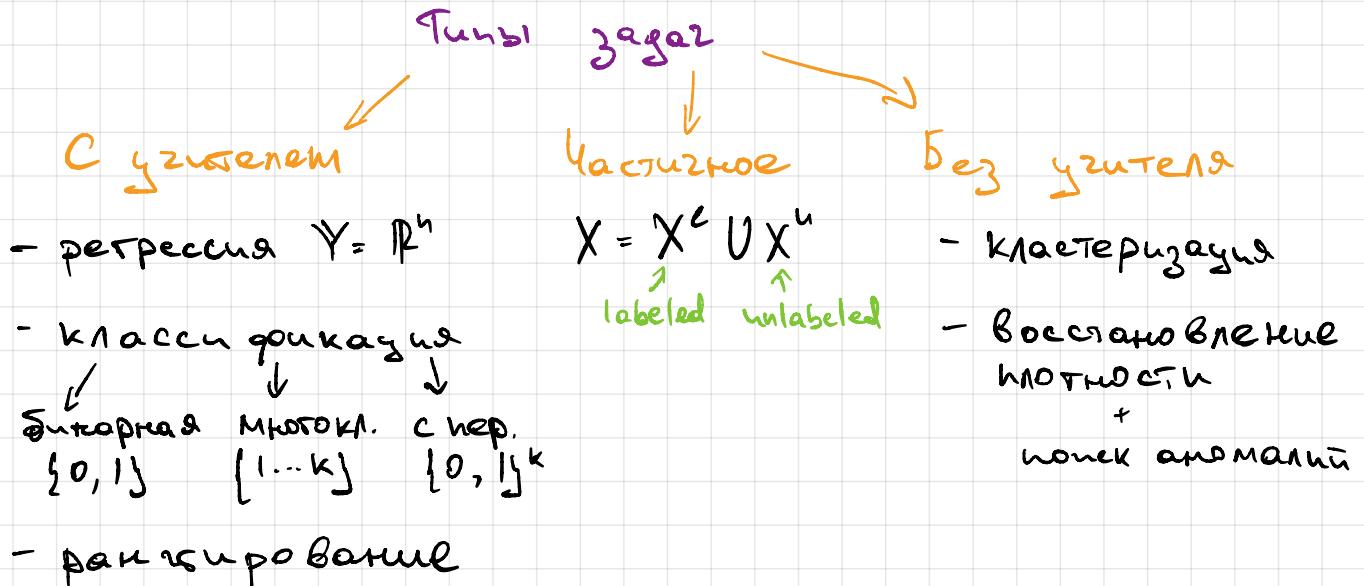




## Семинар 1, 12.09.2023



## Семинар 2, 19.09.2023

Лин. пер.

! Сколько нужно данных?  $10k$  - норм, чтобы наладить закономерности

! Размер тестовой выборки?

- 1) тестовая - перегружательна
- 2) подготовка большая

! Можно строить распределение целевого  $\rightarrow$  гистограммы  
 $\rightarrow$  выборки

! Наго откладывать построение МН модели до определения



Sklearn:

- ① Linear Regression :  $\hat{y}^* = (X^T X)^{-1} X^T y$
- ② Ridge : MSE +  $L_2$ -рег.
- ③ Lasso : MSE +  $L_1$ -рег. - Затягивает признаки

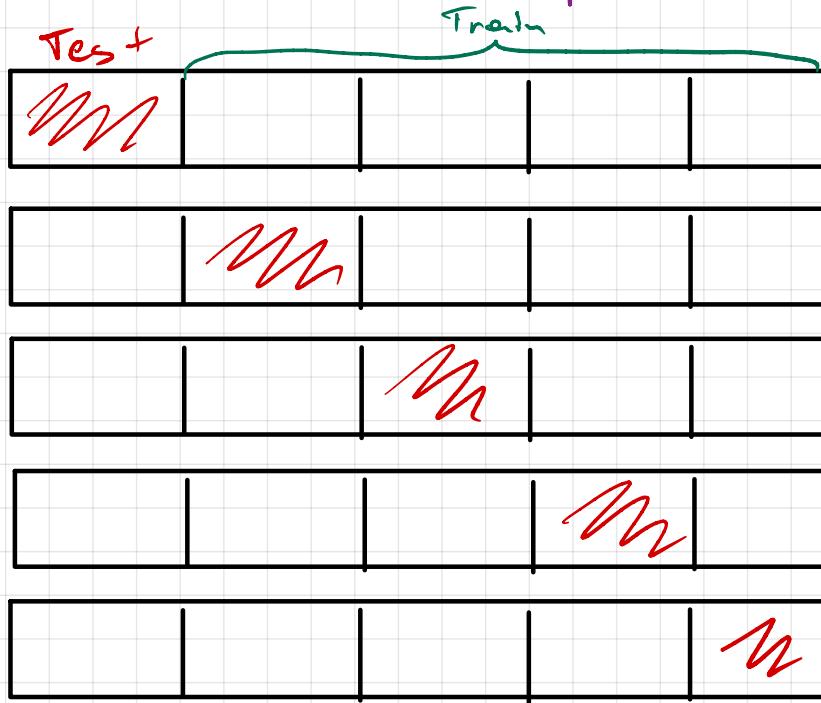
!  $\text{RMSE}_{\text{test}} < \text{RMSE}_{\text{train}}$  ?

↓  
6 + третие  
выборки

↓  
тест  
смешанный

← мало данных

## Кросс-валидация



cross\_val\_score

! Масштаб признаков может сильно отличаться

↓

Веса будут "некорректными"

↓

StandardScaler

Гиперпараметры

Подбираются по GridSearch на валидационной выборке

! Можно строить гистограмму отклонений от таргета

! Можно по признаку разбить объекты на группы и обучить модель для каждой:

"Квартиры с OverallQual > 6 стоят от 200к - обустроены на дорогих и качественных"

Семинар 3 . 26.05.2023

### Матричное дифференцирование.

$$f(x+dx) = f(x) + f'(x)dx + \bar{o}(dx)$$

$$f(x+dx) = f(x) + L(dx) \leftarrow \bar{o}(|dx|)$$

<del>Зависимое Аргумент</del>	Скаляр	Вектор	Матрица
Скаляр	$f'(x)dx$	$\mathbb{J}dx$	-
Вектор	$\nabla f(x)^T dx$	$\mathbb{J}dx$	-
Матрица	$\text{tr}(\nabla f(x)^T dx)$	-	-

$$\textcircled{1} \quad f(x) = \langle a, x \rangle$$

$$f(x+dx) = \langle a, x+dx \rangle = \langle a, x \rangle + \langle a, dx \rangle$$

$$\nabla f(x) = a$$

$$\textcircled{2} \quad f(x) = x^T A x$$

$$\begin{aligned} d(\lambda A + \beta B) &= \lambda dA + \beta dB \\ d(AB) &= (dA)B + A dB \end{aligned}$$

$$\begin{aligned} d(x^T A x) &= d(x^T) A x + x^T d(A x) = \\ &= d(x^T) A x + x^T \left( \underbrace{\frac{d(A)}{0}}_{O} x + A dx \right) = \end{aligned}$$

$$\begin{aligned} d(x^T) A x + x^T A dx &= x^T A^T dx + x^T A dx = x^T (A^T + A) dx = \\ &= \langle (A + A^T)x, dx \rangle \end{aligned}$$

$$\nabla f(x) = (A + A^T)x$$

$$\textcircled{3} \quad f(X) = a^T X A \times a$$

$$\begin{aligned}
 d((a^T)(X A \times a)) &= d(a^T) X A \times a + a^T d((X A \times a)) = \\
 &= a^T d(X A X) a = a^T (d(X) \cdot A X + X \cdot d(A X)) a = \\
 &= a^T dX \cdot A X a + a^T X \cdot A \cdot dX \cdot a = \\
 &= \operatorname{tr}(a^T dX \cdot A X a) + \operatorname{tr}(a^T X \cdot A \cdot dX \cdot a) = \\
 &= \operatorname{tr}((A X a a^T + a a^T X A) dX)
 \end{aligned}$$

$$\nabla f(x) = a a^T X^T A^T + A^T X^T a a^T$$

$$\textcircled{4} \quad f(x) = x^T x \quad \mathbb{R}^n \rightarrow \mathbb{R}^n$$

$$\nabla_x f(x) = ?$$

$$\begin{aligned}
 d(x^T x) &= (dx) x^T x + x d(x^T x) = \\
 &= x^T x dx + 2 x x^T dx = \underbrace{(x^T x \cdot I + 2 x x^T)}_{\sim} dx
 \end{aligned}$$

$$\textcircled{5} \quad \nabla_x \det(X)$$

$$\det X = \sum_{k=1}^n (-1)^{i+k} \cdot X_{ik} M_{ik}$$

↑

$\det X$  degree  $i$ -th column in  $k$ -th row

$$\begin{aligned}
 \frac{\partial}{\partial x_{ij}} \det X &= \sum_{k=1}^n \sum_{i=1}^n (-1)^{i+k} X_{ik} \underbrace{M_{ik}}_{\text{const}} = \sum_{j=1}^n (-1)^{i+j} X_{ij} M_{ij} = \\
 &= (-1)^{i+j} M_{ij}
 \end{aligned}$$

$$\left( \frac{\partial \det X}{\partial x_{ij}} \right)_{i,j=1}^n = \begin{bmatrix} M_{11} & M_{12} & \dots \\ \vdots & & \end{bmatrix}$$

$$(X^{-1})_{ij} = \frac{1}{\det X} (-1)^{i+j} M_{ji}$$

$$(X^{-T})_{ij} = \frac{1}{\det X} (-1)^{i+j} M_{ij}$$

$$(-1)^{i+j} M_{ij} = \det X \quad (X^{-T})_{ij} = \frac{1}{\det X} \det X$$

$$\forall \det X: (\det X) X^{-T} \Rightarrow df = \text{Tr}(X^{-1} \det X \cdot dX)$$

$$\begin{aligned} \textcircled{6} \quad d_A \text{tr}(AB) &= \text{tr}(d(AB)) = \text{tr}((dA)B + A dB) = \\ &= \text{tr}(B dA) = \text{tr}(\nabla f(A)^T dA) \end{aligned}$$

$$\nabla f(A) = B^T$$

$$\textcircled{7} \quad \nabla_A \text{tr}(x^T A y)$$

$$d_A \text{tr}(x^T A y) = \text{tr}(d_A(x^T A y)) = \text{tr}(x^T d(Ay)) =$$

$$= \text{tr}(x^T dA \cdot y) = \text{tr}(y x^T dA)$$

$$\nabla f(A) = xy^T$$

$$\textcircled{8} \quad \nabla_x g(\underbrace{f(x)}_{\text{f}(x)}) = g'(f(x)) \nabla_x f(x)$$

$$\nabla_w \log(1 + e^{-y \langle \omega, x \rangle}) = \nabla_w \underbrace{\log(1 + e^{-y \omega^T x})}_{g(f(\omega))} =$$

$$= g'(f(\omega)) \cdot \nabla_\omega f(\omega) = \frac{1}{1 + e^{-y \omega^T x}} \cdot e^{-y \omega^T x} \cdot (-y) \cdot \nabla_\omega f(\omega) = \dots$$

$$df(\omega) = d(\omega^T x) = d(x^T \omega) = x^T d\omega \Rightarrow \nabla_\omega f(\omega) = x$$

$$\dots = \left[ \zeta(z) = \frac{1}{1 + e^{-z}} \right] = - \langle (-y \langle \omega, x \rangle) \cdot e^{-y \langle \omega, x \rangle}, yx \rangle$$

$$\textcircled{2} \quad \|X\omega - y\|_2^2 \rightarrow \min$$

MSE

$\hat{f}(\omega)$

$$f(\omega) = \|X\omega - y\|_2^2 = (X\omega - y)^T (X\omega - y) = g(\underbrace{X\omega - y}_{\tilde{f}(\omega)}),$$

$$\text{тогда } g(z) = z^T z$$

$$\tilde{f}: \mathbb{R}^d \rightarrow \mathbb{R}^e, \quad g: \mathbb{R}^e \rightarrow \mathbb{R}$$

$$\nabla_{\omega} f = \tilde{J}_{\tilde{f}}^T(\omega) \cdot \nabla_z g(z)|_{z=\tilde{f}(\omega)}$$

$$\nabla_z g(z) = 2z$$

$$d\tilde{f}(\omega) = d(X\omega - y) = \underbrace{X}_{\text{d}\omega} d\omega \Rightarrow \tilde{J}_{\tilde{f}}(\omega) = X$$

$$\Rightarrow \nabla_{\omega} f = X^T \cdot 2(X\omega - y) = 2X^T(X\omega - y)$$

$$df = 2(X\omega - y)^T X d\omega, \text{ но иначе вид}$$

$$\text{Мат. ГД: } \omega_t = \omega_{t-1} - \gamma \cdot 2X^T(X\omega_{t-1} - y)$$

Семинар 4. 03.10.2023

Подготовка данных

11) Вывод ошибки - обзоры, отражающиеся от большинства

a) Некорректные данные

б) Редкие данные

в) Не хватает признаков для обозначения целевой переменной

Как искать: - вычленяции

⇒ поиск аномалий (В МО-2)

Что делать: - выкидывать

- робастные методы (менее чувствительны к выбросам - MAE)

- не трогать их (рек. системы)

## ② Пропуски (missing data)

- Что делать:
- выбрасывать объекты/признаки
  - заполнение (среднее, сим. значение)  
может
  - ↓ + сделать признак "был ли пропуск"

## ③ Категориальные признаки

### 3.1 Label encoding

Басмачий  $\rightarrow 1$

Задаем порядок,  
которого нет

Хановка  $\rightarrow 2$

Бургово  $\rightarrow 3$

:

### 3.2 OHE

### 3.3 Mean-target-encoding (MTE)

район	y	MTE
X	100	80
Б	50	50
Н	15	16
Н	17	16
X	80	80

Почему это опасно? Утечка целевой переменной  
(target leak)

Почему это? 1) Если много редких категорий, то на существующей gone выборке  $MTE = y$

2) MTE - простая модель  $b(x)$

Обучена на обуз. выборке

на обуз. выборке коечко пытается

итого ват

модель  $\rightarrow q(x, b(x)) \rightarrow$  на обуз. | на тесте

$b(x)$  - см. е.

$b(x)$  - см. як

увенч.

$a(x)$  рассчитывается  
 на основе  
 (рассчитывается  
 на такое расп.  
 омисок)

$a(x)$  рассчитывается  
 на основе  
 (рассчитывается  
 на старое)

другое расп.  
 омиски  
 ↓  
 $a(x)$  всё еще  
 рассчитывается  
 на старое

Как бороться? - добавить шум:  $MTE(x) + \epsilon$

$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

Гиперпараметр

Семинар 5, 10.10.2023

## Mean Target Encoding

$f_j(x)$  - j-й признак катер.

Бин. класс.:

$$g_j(x, X) = \frac{\sum_{i=1}^n [f_j(x_i) = f_j(x)] \cdot [y_i = +1]}{\sum_{i=1}^n [f_j(x_i) = f_j(x)]} -$$

где хотим спреди всех с таким  
значением

Многокласс.:

$$g_{jk}(x, X) = \frac{\sum_{i=1}^n [f_j(x_i) = f_j(x)] \cdot [y_i = k]}{\sum_{i=1}^n [f_j(x_i) = f_j(x)]} -$$

$k = 1 \dots K$  - классы

где k-го класса

Перекрестка:

$$g_j(x, X) = \frac{\sum_{i=1}^n [f_j(x_i) = f_j(x)] \cdot y_i}{\sum_{i=1}^n [f_j(x_i) = f_j(x)]} -$$

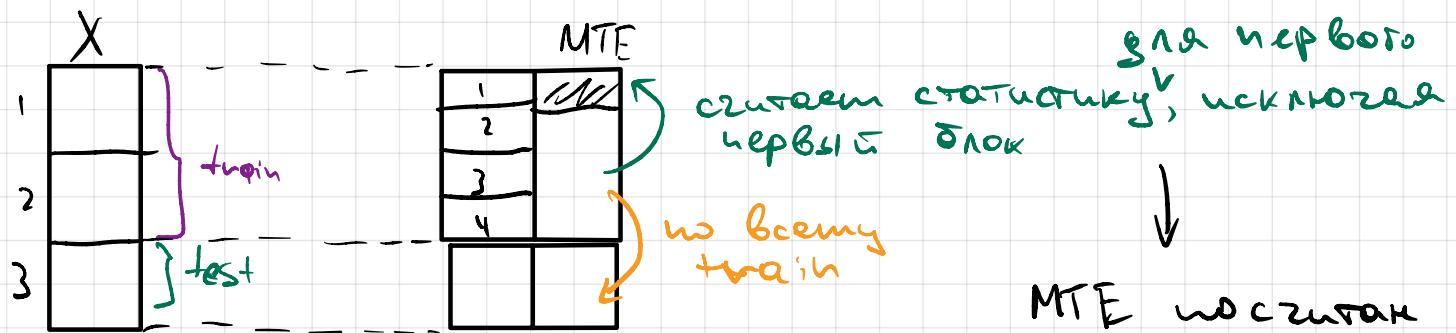
Проблема: data leak

	MTE	X	y
train			
test			

MTE - простая модель  $\Rightarrow$  можно  
 одержать навигу, это в  
 тесте он несет неоптимальные

# Как бороться?

## ① Кросс-валидация



для первого  
MTE исключают  
как будто для  
тестовой выборки

## ② Зашумление

$$g_i(x, X) + \varepsilon$$

$\sim \mathcal{N}(0, \sigma^2)$

гиперпараметр

## ③ Сглаживание

$$g_j(x, X) = \hat{\pi}(n_j) \frac{\sum_{i=1}^l [f_j(x_i) = f_j(x)] \cdot [y_i = +1]}{\sum [f_j(x_i) = f_j(x)]} + (1 - \hat{\pi}(n_j)) \frac{1}{l} \sum_{i=1}^l [y_i = +1]$$

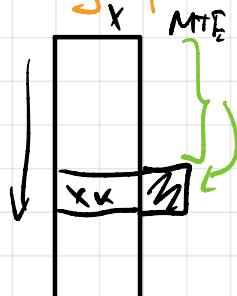
$$n_j = \sum_{i=1}^l [f_j(x_i) = c_j]$$

На все можно не ставить формуле

$$\hat{\pi}(n) = \frac{1}{1 + e^{-dn - \beta}} \quad \begin{matrix} \uparrow \\ \text{гиперпараметры} \end{matrix}$$

Больше учитывается, если  $n_j \gg 0$

## ④ Копирование во времени (Copy Boost)



$$g_j(x, X) = \frac{\sum_{i=1}^{k-1} \dots}{\sum_{i=1}^k \dots}$$

Улучшение: посчитать несколько MTE, каждый для своего порядка

Если времена в разных нет, то сортируем рандомно

В тесте считаем зерчило во всему Test

А что это?

## ① Counters

$$g_j(x, X) = \sum_{i=1}^l [f_j(x_i) = f_j(x)]$$

## ② Weight of Evidence (Доказательство)

$$P(C | y=b) = \frac{\sum_{i=1}^l [y_i = b] [f_j(x_i) = C] + \alpha}{\sum_{i=1}^l [y_i = b] + 2\alpha}$$

Категория

- всех объектов  
категории C в  
классе b

$$g_j(x, X) = \log \left( \frac{P(f_j(x) | y=+1)}{P(f_j(x) | y=-1)} \right)$$

объектов этого кат.  
доминантные среди +1 и -1  
класса

$$L \in N$$

## Работа с текстами

## ① Bag of words

Словарь:  $\vartheta_1, \dots, \vartheta_m$

и наборов признаков:  $g_j$  - число  $\vartheta_x$ .  $\vartheta_i$  в текст

## ② TF-IDF

$$tf(t, d) = \frac{\# \text{слова } t \text{ в док. } d}{\# \text{слов в } d}$$

частота слова в тексте

$$idf(t, D) = \log \frac{|D|}{|\{d \in D | t \in d\}|}$$

одинокое слово

для каждого слова:  $tf \cdot idf$

слова много в тексте  $\Rightarrow \uparrow$   
 слова мало в многих текстах  $\Rightarrow \downarrow$

## ③ Пред обработка

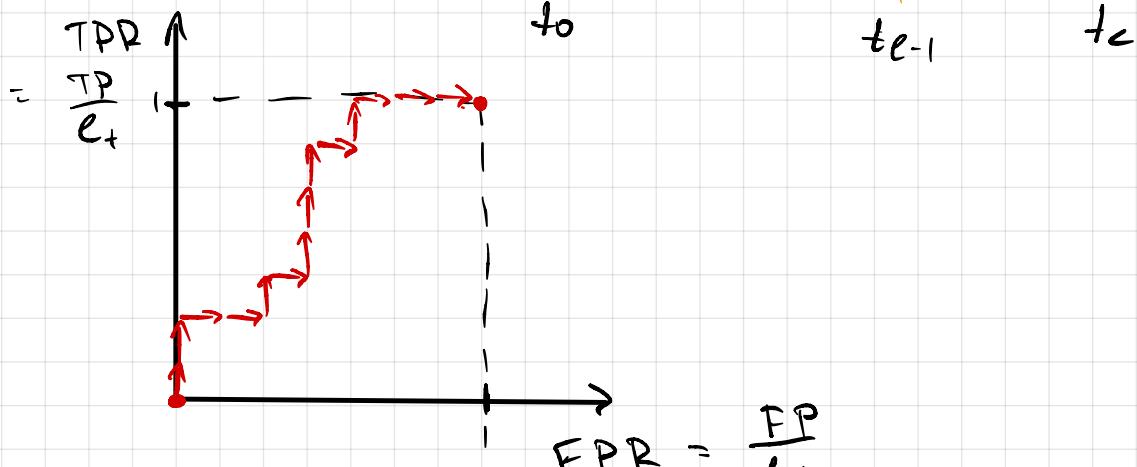
- очистка
- лемматизация
- токенизация
- n-граммами

## ROC-кривая

⑥ Анализуемая оценка AUC-ROC

$b(x)$  — бывает уверенностью 0 или 270  $y(x) = +1$

Отсортируем:  $b(x_{(1)}) < \dots < b(x_{(L)})$



$$(a(x) = +1, y = -1)$$

TPR

$$t_e : 0 \quad 0$$

$$t_{e-1} : y(e) = +1 \quad +\frac{1}{L+1} \quad +0$$

$$y(c) = -1 \quad +0 \quad +\frac{1}{L}$$

Зеро  $y = +1 \rightarrow$  Блеск

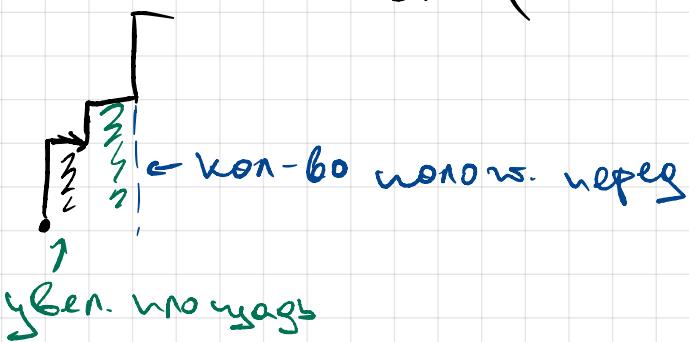
Зеро  $y = -1 \rightarrow$  Буря

$$t_0 : 1 \quad 1$$

$$AUC-ROC = 1 \Leftrightarrow +1 \dots +1 -1 \dots -1$$

No генетики!: +1 -1 +1 -1 ... +1 -1

$$AUC-ROC = \frac{1}{C \cdot C} \sum_{i=1}^C \left( [y_{(i)} = -1] \cdot \sum_{j=i+1}^C [y_{(j)} = +1] \right) - \text{не залогоритм}$$



① Наименование AUC-ROC с зонами неизвестных нап

$$AUC-ROC = \frac{1}{C \cdot C} \sum_{i=1}^C \sum_{j=i+1}^C [y_{(i)} = -1] \cdot [y_{(j)} = +1] = \dots$$

зарифметичное  
направление

$\begin{matrix} -1 & +1 \\ i & j \end{matrix}$

→  $b(x)$

$$\dots = \frac{1}{C \cdot C} \sum_{i < j} \left( 1 - [y_{(i)} = y_{(j)}] - [y_{(i)} > y_{(j)}] \right) =$$

$$= \frac{1}{C \cdot C} \sum_{i < j} \left( \underbrace{1 - [y_{(i)} = y_{(j)}]}_{[y_{(i)} \neq y_{(j)}]} \right) - \frac{1}{C \cdot C} \sum_{i < j} [y_{(i)} > y_{(j)}] =$$

$$= \frac{C+C}{C \cdot C} - \frac{1}{C \cdot C} \sum_{i < j} [y_{(i)} > y_{(j)}] = \dots$$

$$\text{Доля зерп. напр.} = \frac{2}{C(C-1)} \cdot \sum_{i < j} [y_{(i)} > y_{(j)}]$$

$$DP = \frac{2C}{C(C-1)} (1 - AUC-ROC)$$

②  $a(x)$

$$AUC-ROC(a) < 0,5$$

Как сделать  $> 0,5$ ?

$$\tilde{a}(x) = -a(x)$$

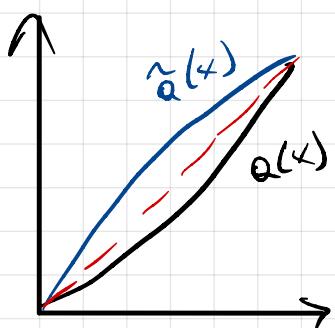
$$\ell_- \cup \ell_+ = \text{const}$$

$$TP(\hat{\alpha}) = \sum_{i=1}^{\ell} [y_i = +1] \cdot [\hat{\alpha}(x_i) = +1] = \sum_{\{y_i = +1\}} [\alpha(x_i) = +1] \geq FN(\alpha)$$

$$FP(\hat{\alpha}) = \sum_{i=1}^{\ell} [y_i = -1] \cdot [\hat{\alpha}(x_i) = +1] = \sum_{\{\alpha(x_i) = -1\}} [\hat{\alpha}(x_i) = +1]$$

$$TPR(\hat{\alpha}) = \frac{TP(\hat{\alpha})}{\ell_+} = \frac{FN(\alpha)}{\ell_+} = \frac{\ell_+ - TP(\alpha)}{\ell_+} = 1 - FPR(\alpha)$$

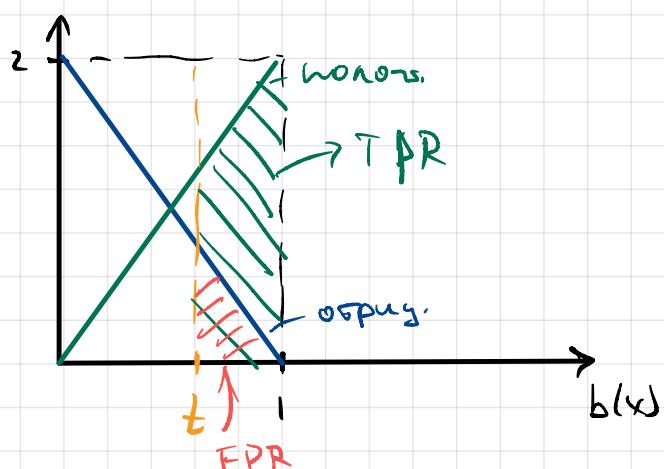
$$FPR(\hat{\alpha}) = 1 - TPR(\alpha)$$



$$\textcircled{3} \quad b(x) \in [0, 1]$$

На  $(0, 1]$  объекты отнесены к классу распределены  $p(b) = 2 - 2b$   
на  $\{0\}$ . класса  $b(x)$   $p(b) = 2b$

ROC - формула и шаги



$$[b(x) \geq t]$$

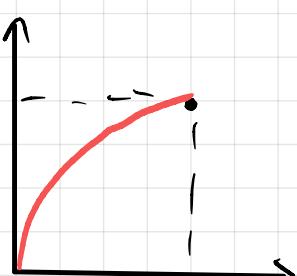
$$TPR(b, +) = 1 - \frac{1}{2}t \cdot 2t = 1 - t^2$$

$$FPR(b, +) = (1-t)^2$$

$$\begin{aligned} TPR(t) &= 1 - (1 - \sqrt{FPR})^2 = \\ &= 1 - 1 + 2\sqrt{FPR} - FPR = \\ &= 2\sqrt{FPR} - FPR \end{aligned}$$

ROC - кривая

$$y = 2\sqrt{x} - x$$



$$AUC - ROC = \int_0^1 (25x - x) dx = \frac{5}{6}$$

④  $\ell = 4000$  - клиенті

1) Барың ғенет  $uQ = 800$  зерткеков

2) Аналитикалық үшасының, 250 откликнаның  $6\%$

3) Ностроуның  $Z$  мөдөнү

$$A: TPR = 0.2, FPR = 0.1$$

$$B: TPR = 0.6, FPR = 0.25$$

$$\alpha(x) = +1 - \text{ клиент күннүр}$$

$$800 = TPR \cdot \ell_+ + FPR \cdot \ell_- \rightarrow \text{хосум}$$

$$\ell_+ = 4000 \cdot 0.06 = 240$$

$$TPR = \frac{10}{3} - \frac{47}{3} FPR$$

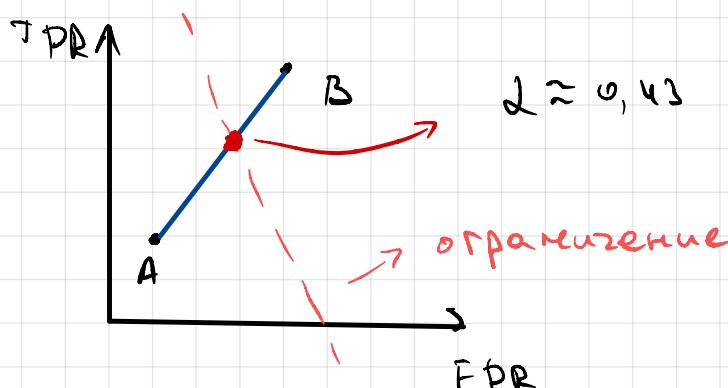
$$\ell_- = 3760$$

$$A: 0.2 \cdot 240 + 0.1 \cdot 3760 = 424$$

$$B: 1084$$

$$C: X \xrightarrow{\alpha} A(x) \\ \xrightarrow{1-\alpha} B(x)$$

$$TPR(C) = \alpha \cdot TPR(A) + (1-\alpha) \cdot TPR(B)$$

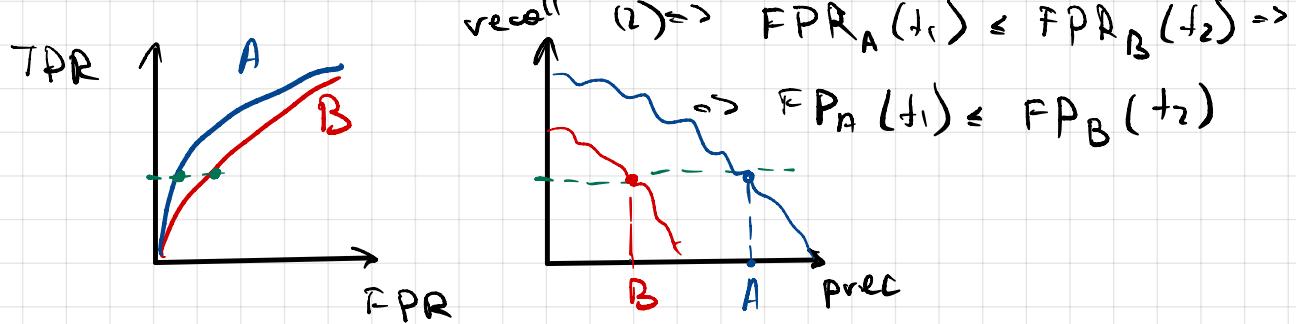


⑤  $\ell_+, \ell_-$  - ғонкырловалы

$$ROC(A) \geq ROC(B) \quad (\text{следует не мене})$$

$$PR(A) \geq PR(B)$$

①  $t_1, t_2$  - wązory :  $\text{TPR}_A(t_1) = \text{TPR}_B(t_2) \Rightarrow (1) \text{TP}_A(t_1) = \text{TP}_B(t_2)$



$$\text{prec}_A(t_1) = \frac{\text{TP}_A(t_1)}{\text{FP}_A(t_1) + \text{TP}_A(t_1)} \geq \frac{\text{TP}_B(t_2)}{\text{FP}_B(t_2) + \text{TP}_B(t_2)} = \text{prec}_B(t_2)$$

② Awanormy