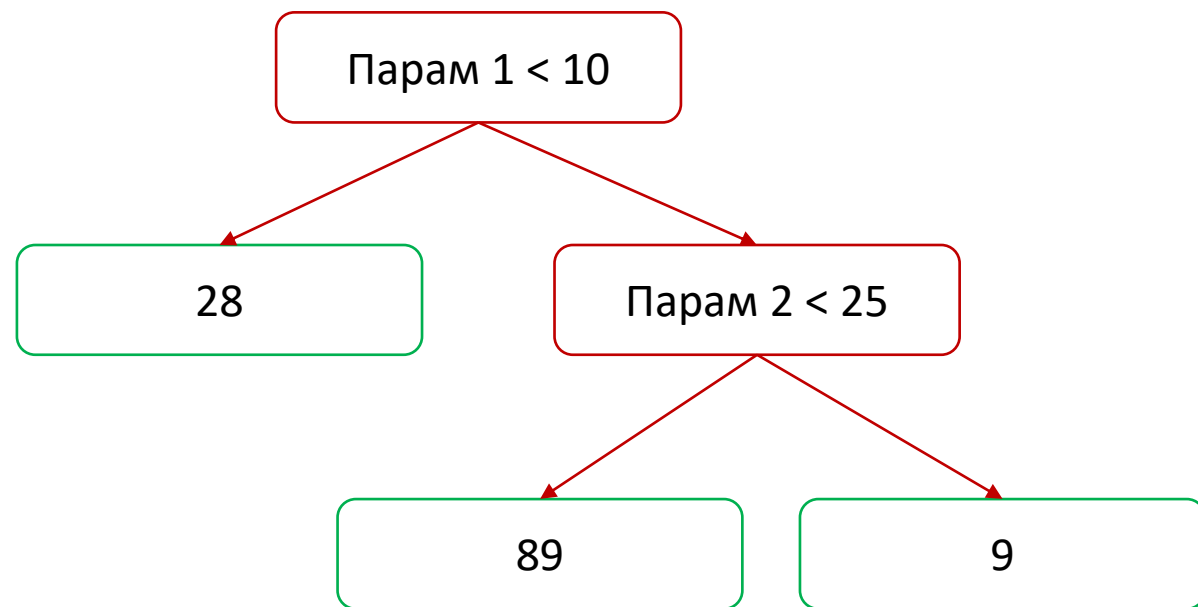


???GBoost under the hood

HEAD OF CVM @ MAGNIT
SELEZNEV A.A.

ПЛАН ПО GBT-РЕГРЕССИИ

1. Интуитивный разбор регрессии на решающих деревьях
(как это в Gboosting?)



ПЛАН ПО GBT-РЕГРЕССИИ

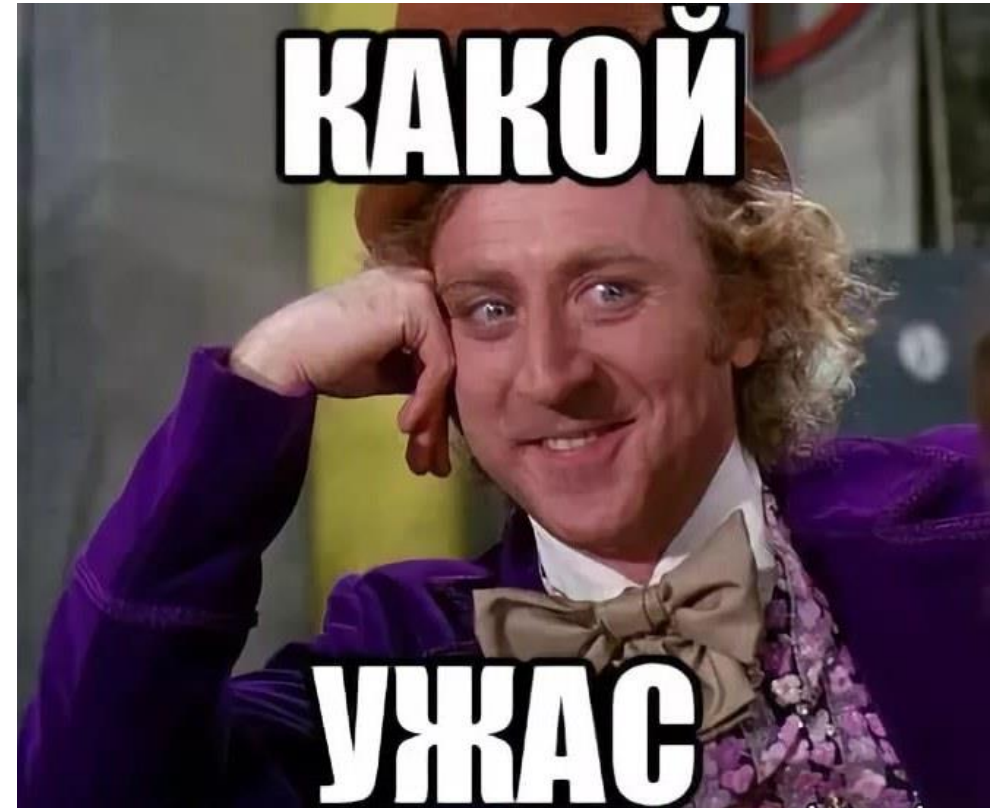
1. Интуитивный разбор регрессии на решающих деревьях
(как это в Gboosting?)
2. Математика под капотом

$$r_{i,m} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$

$$F_m(x) = F_{m-1}(x) + \nu \sum \gamma_m I(x \in R_{jm})$$

ПЛАН ПО GBT-РЕГРЕССИИ

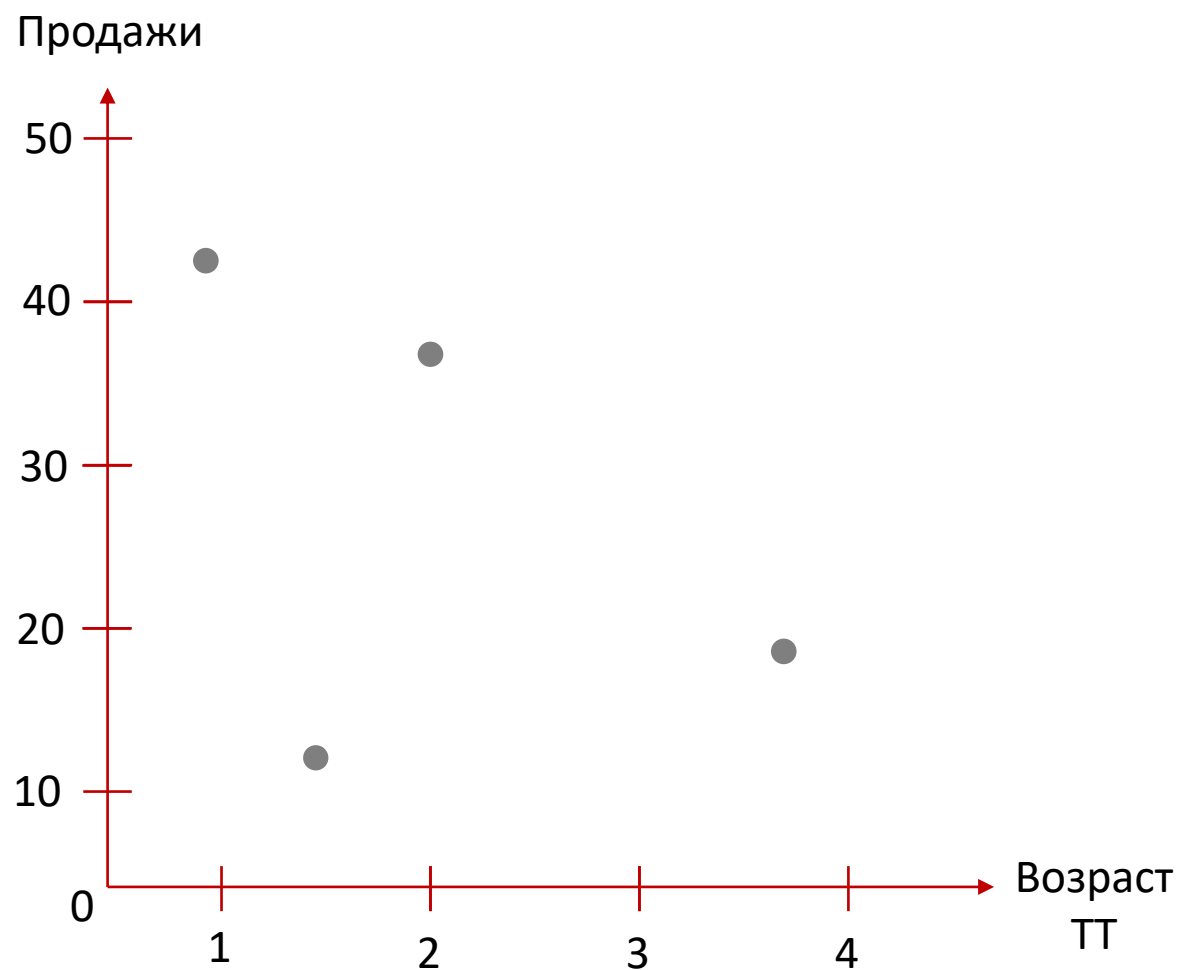
1. Интуитивный разбор регрессии на решающих деревьях
(как это в Gboosting?)
2. Математика под капотом
3. Выбор случайного докладчика на тему
«Энтропия»



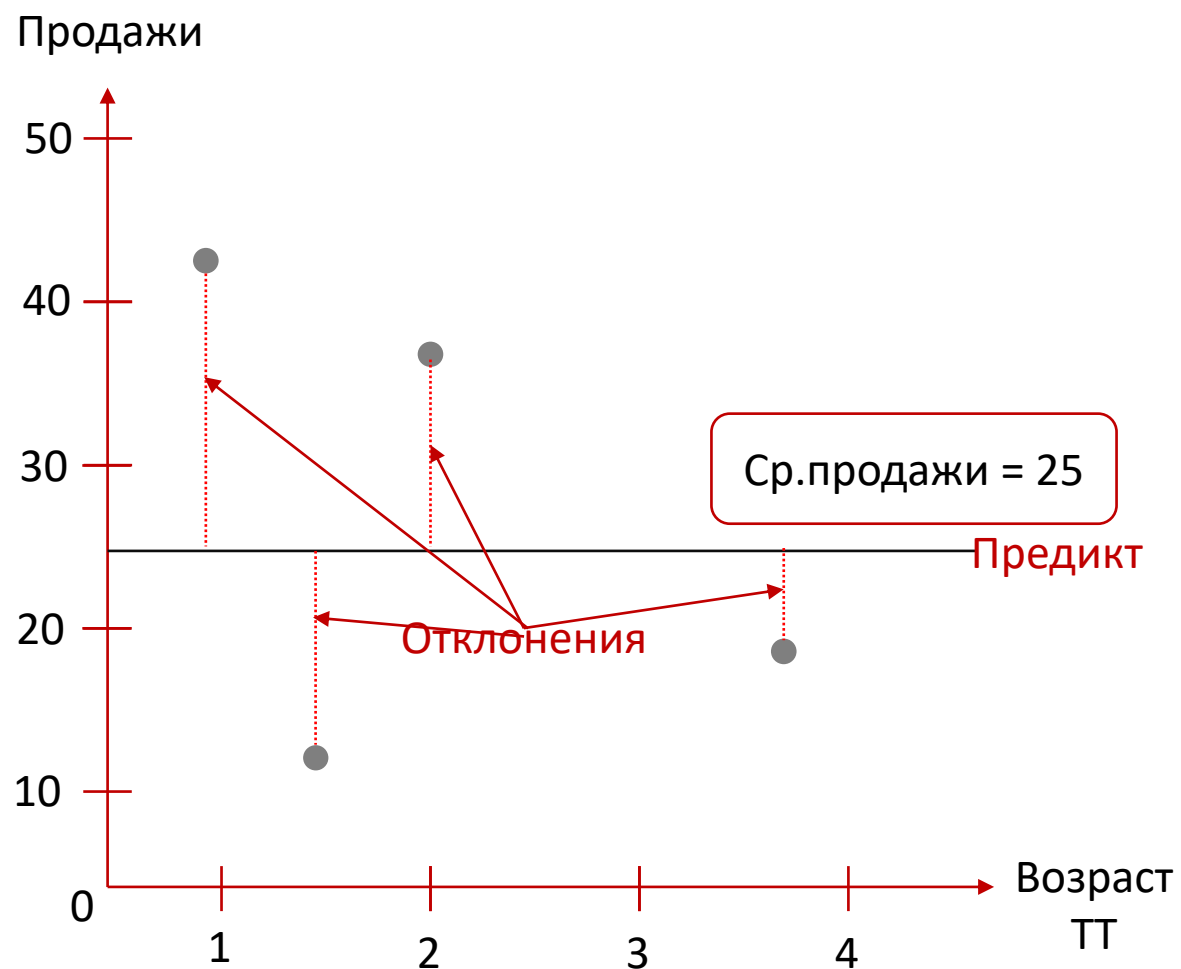
GAIN В ДЕРЕВЬЯХ



GAIN?

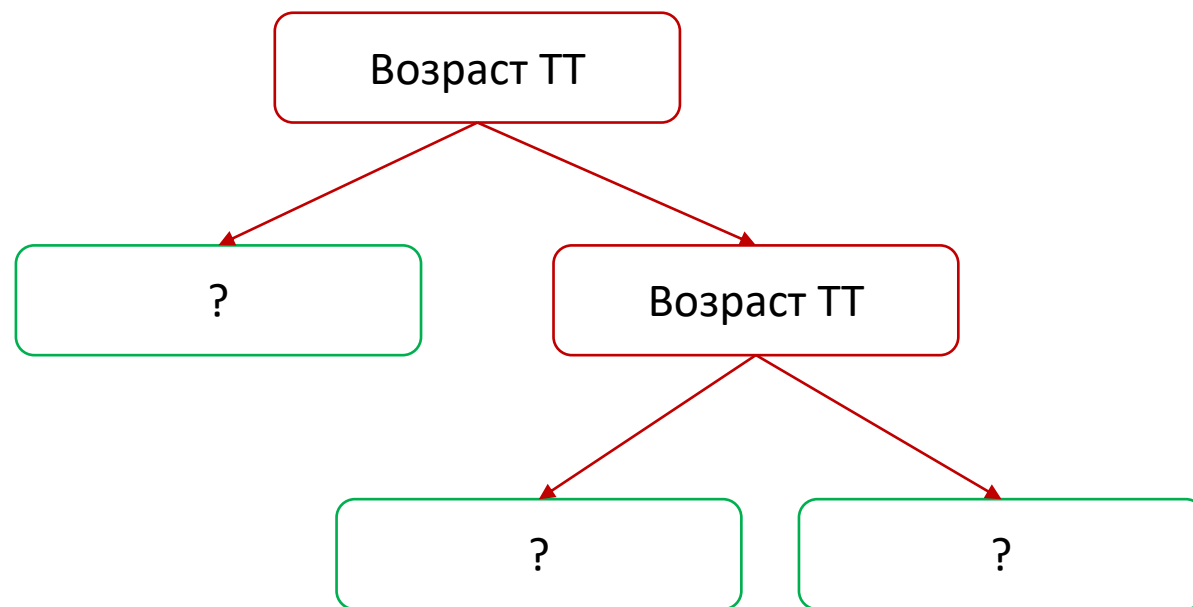
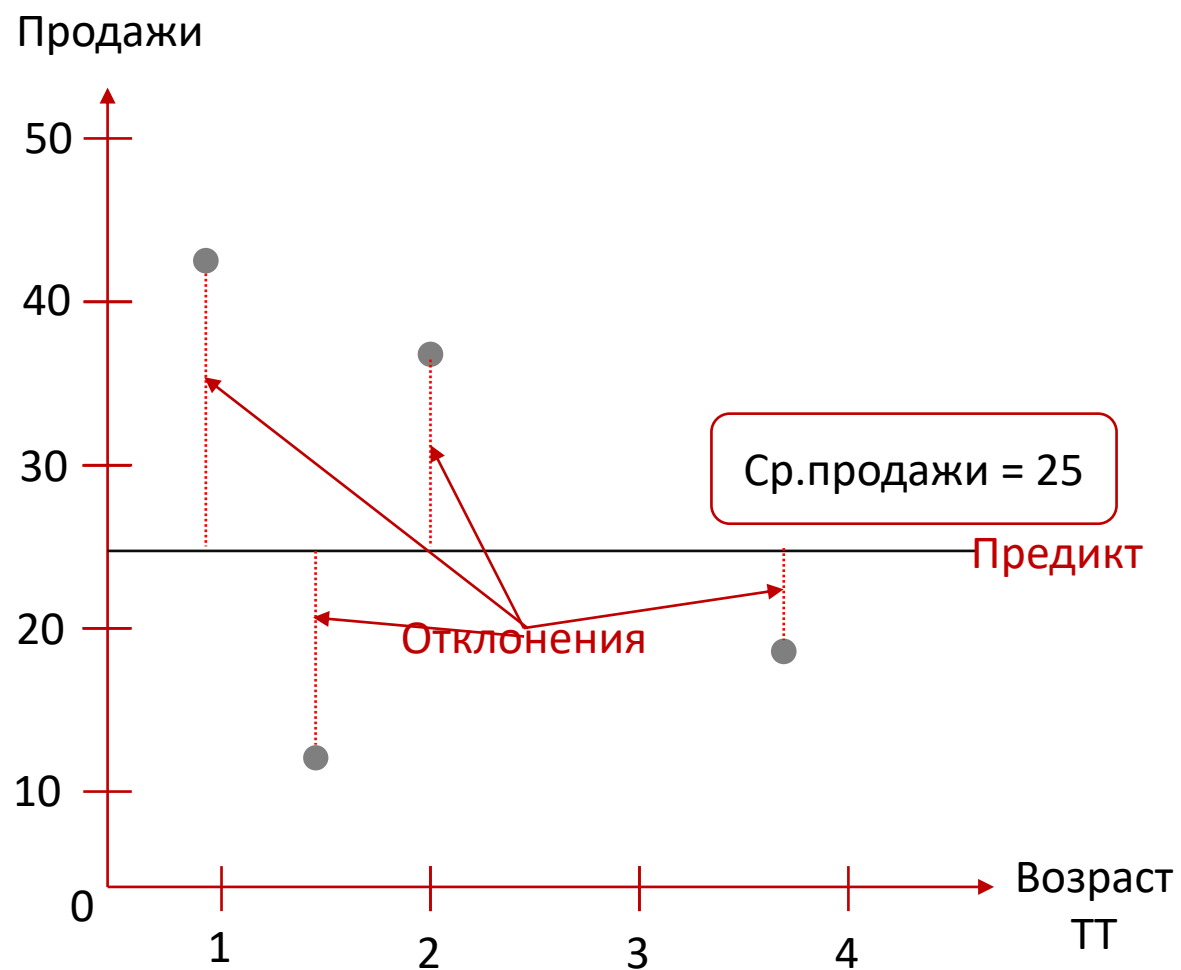


GAIN?

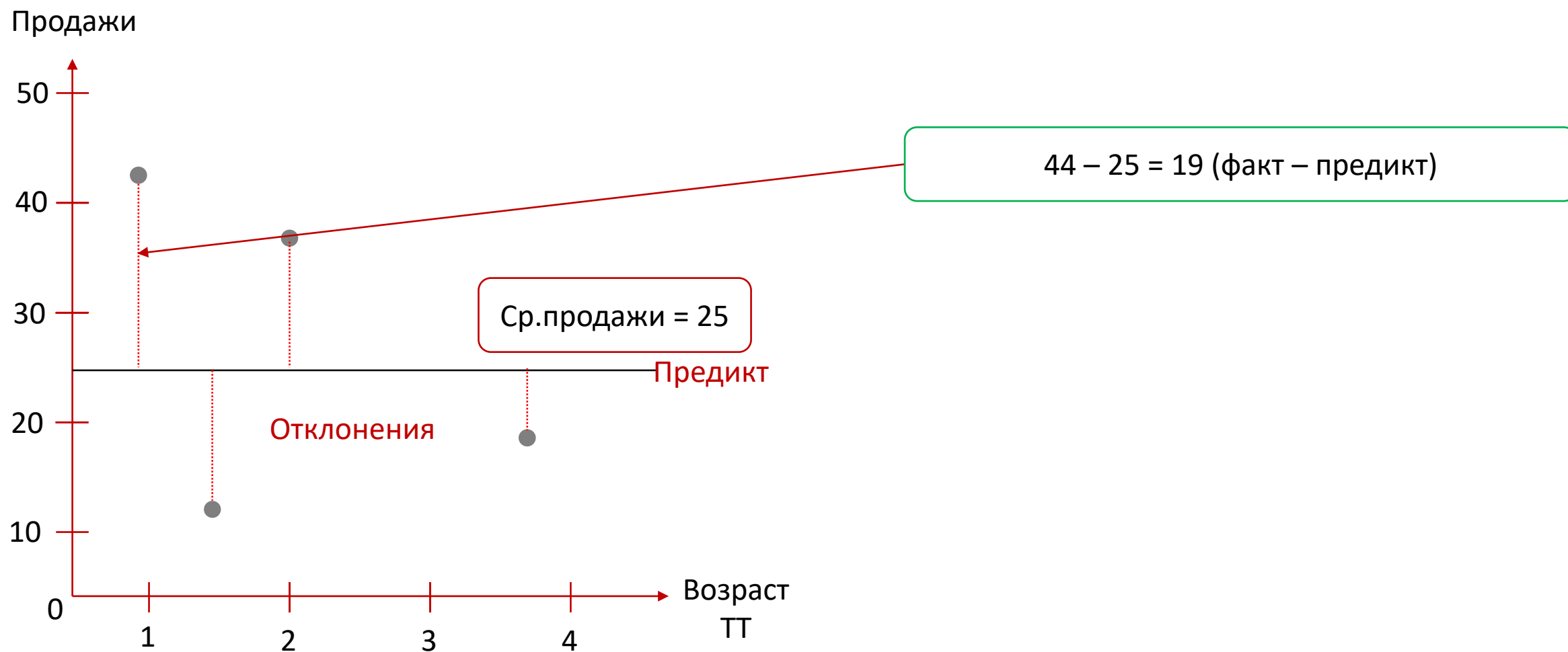


1. Разность – это разница между фактическим значением и его отклонением от предсказанного значения (25)

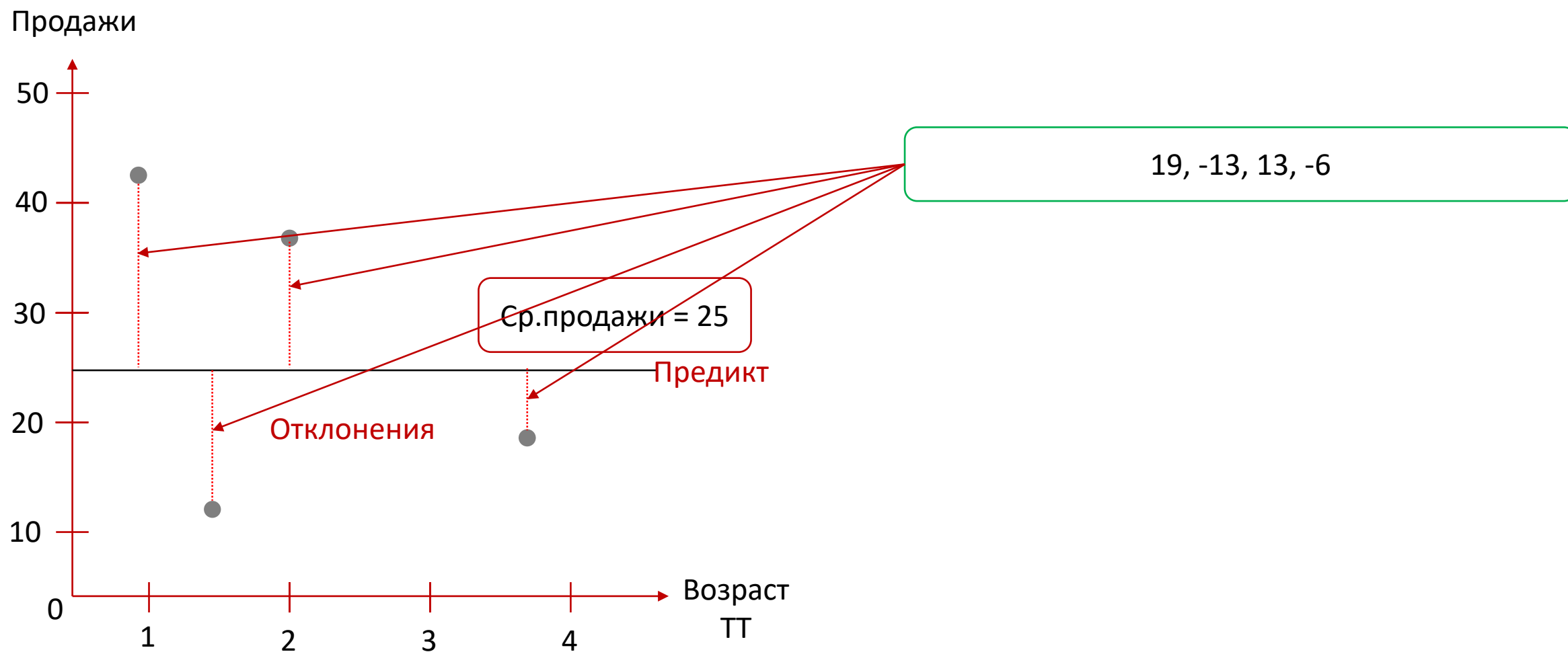
GAIN?



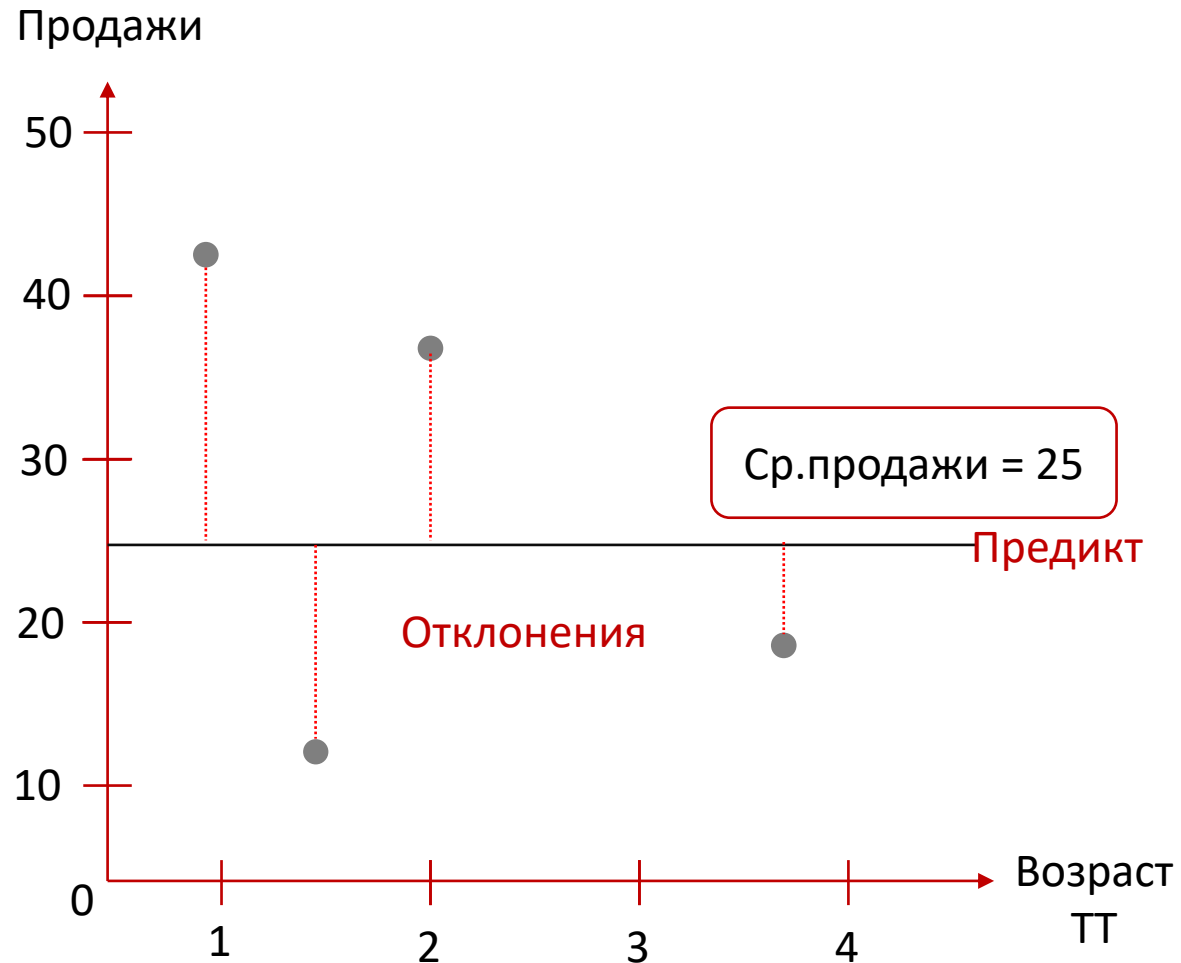
GAIN?



GAIN?



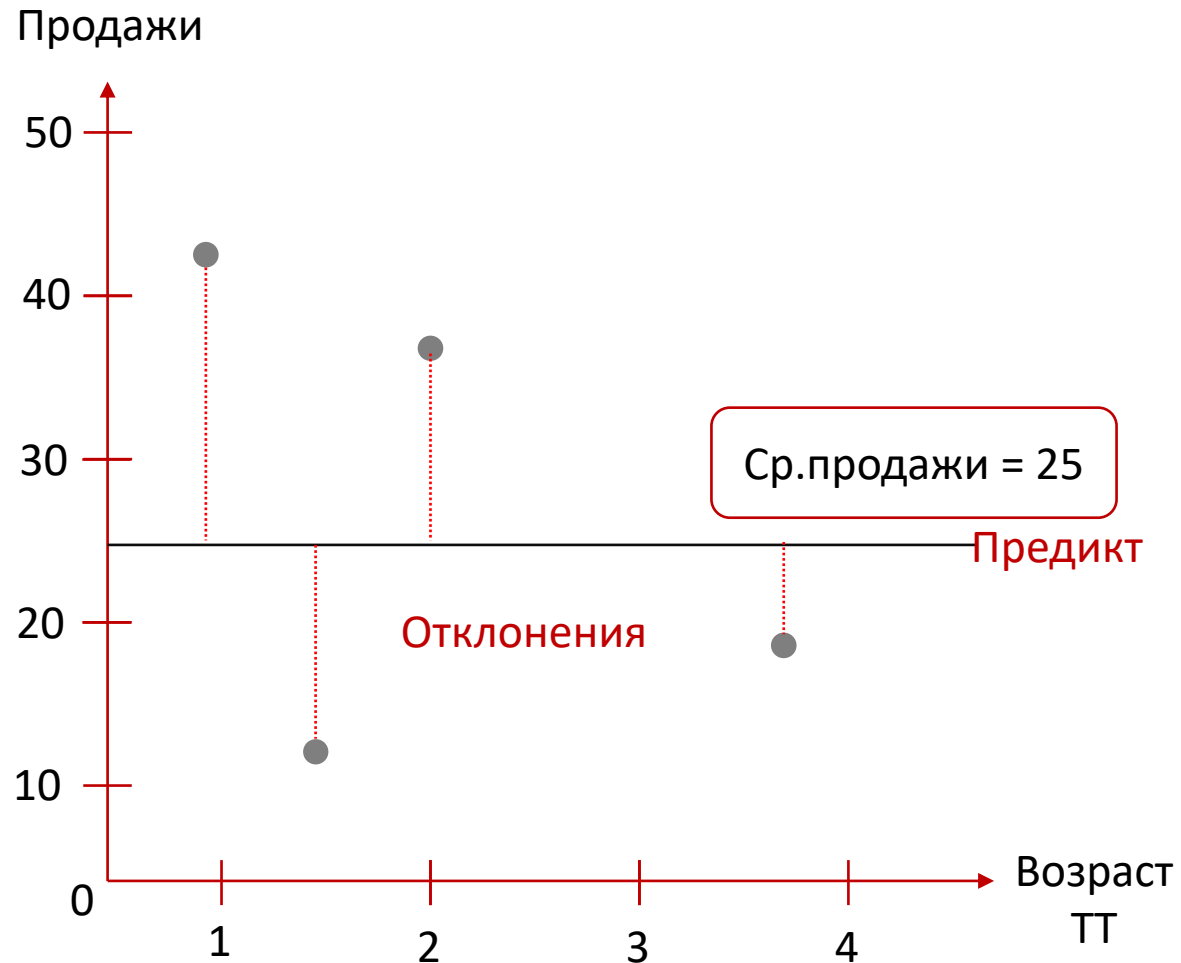
SIMILARITY SCORE



19, -13, 13, -6

$$\text{Скоринг} = \frac{\text{Квадрат суммы разниц}}{\text{Кол-во разниц} + L}$$

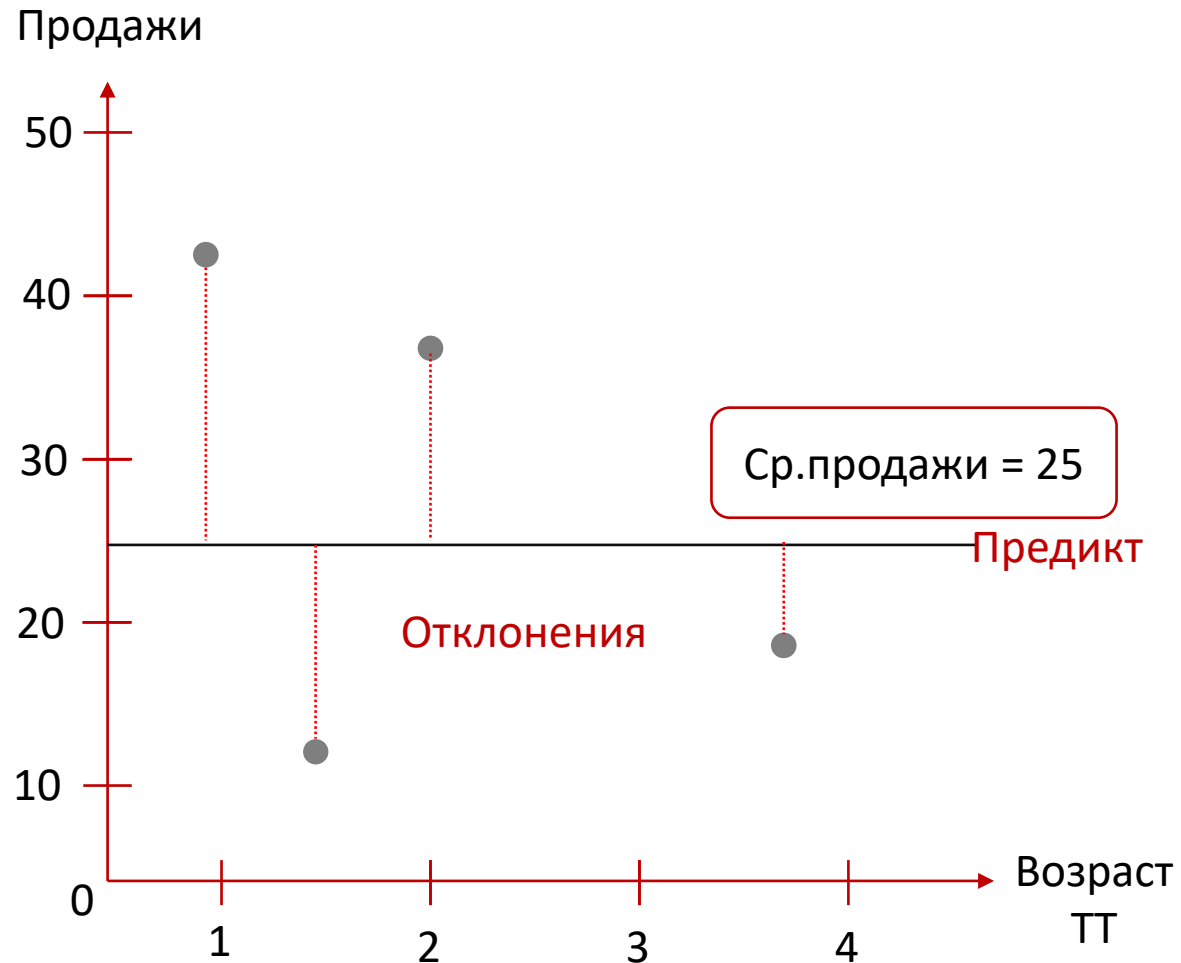
SIMILARITY SCORE



19, -13, 13, -6

L – lambda – параметр защиты от переобучения

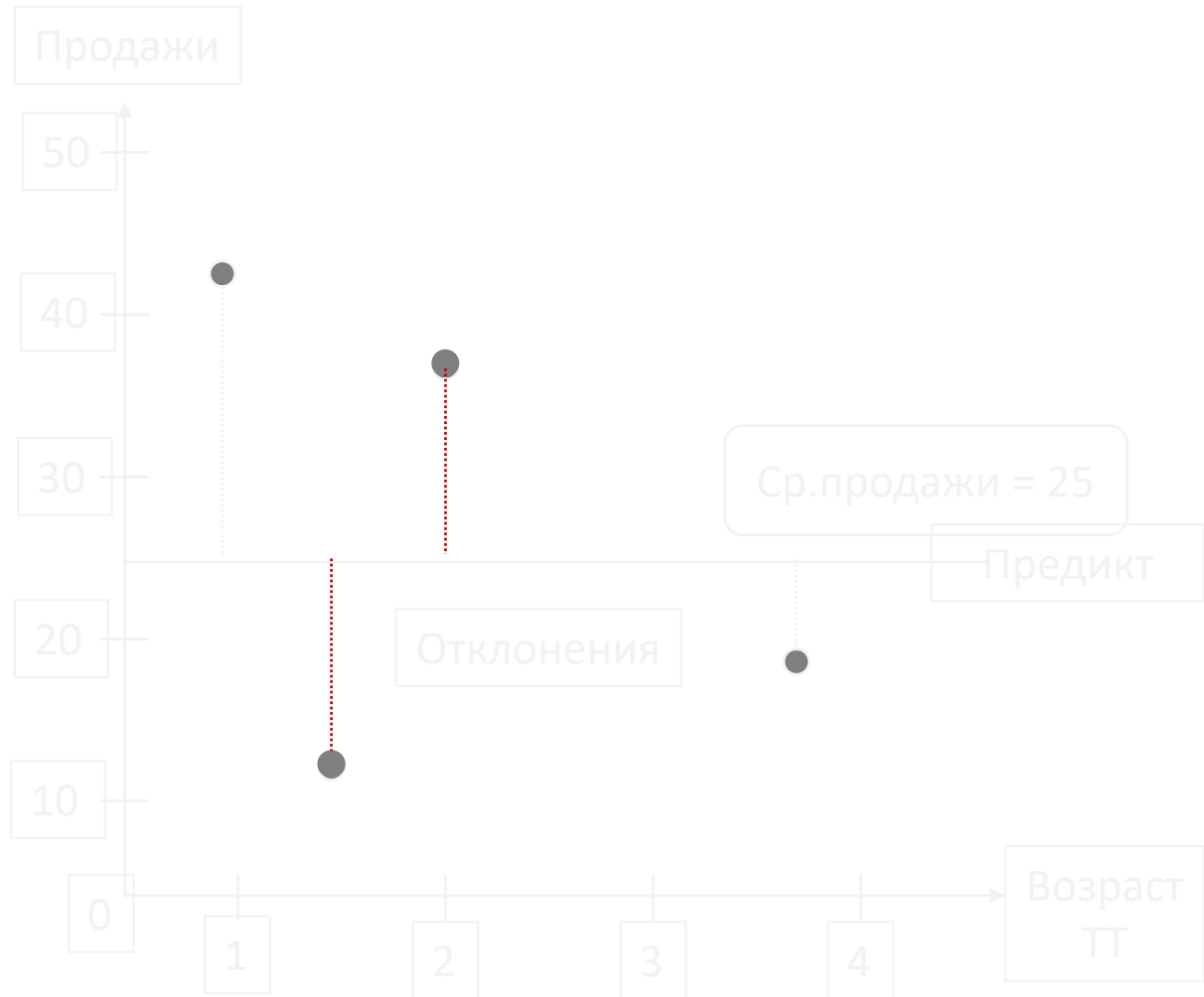
SIMILARITY SCORE



19, -13, 13, -6

$$\text{Скоринг} = \frac{(19 - 13 + 13 - 6)**2}{4 + 0}$$

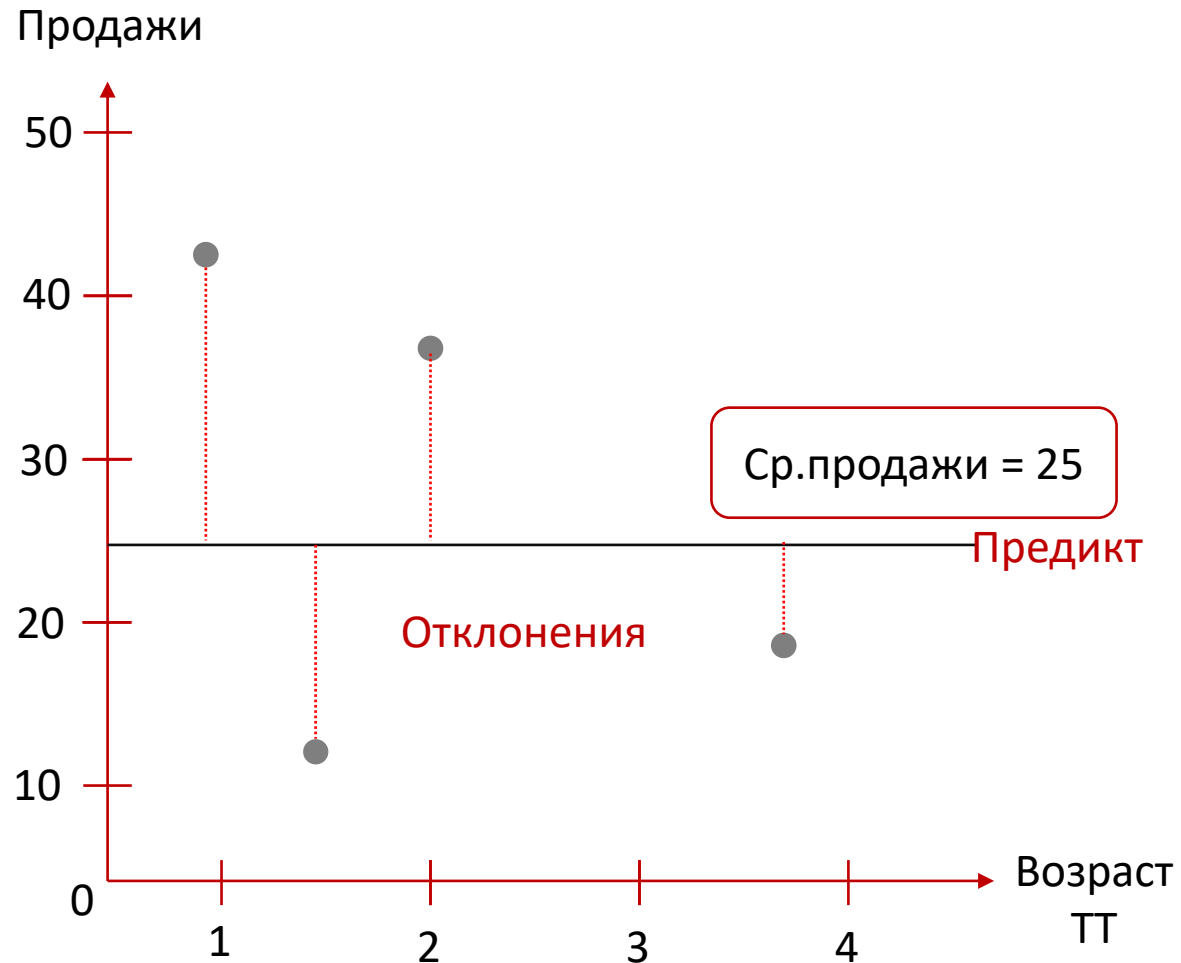
SIMILARITY SCORE



19, -13, 13, -6

$$\text{Скоринг} = \frac{(19 - 13 + 13 - 6)**2}{4 + 0}$$

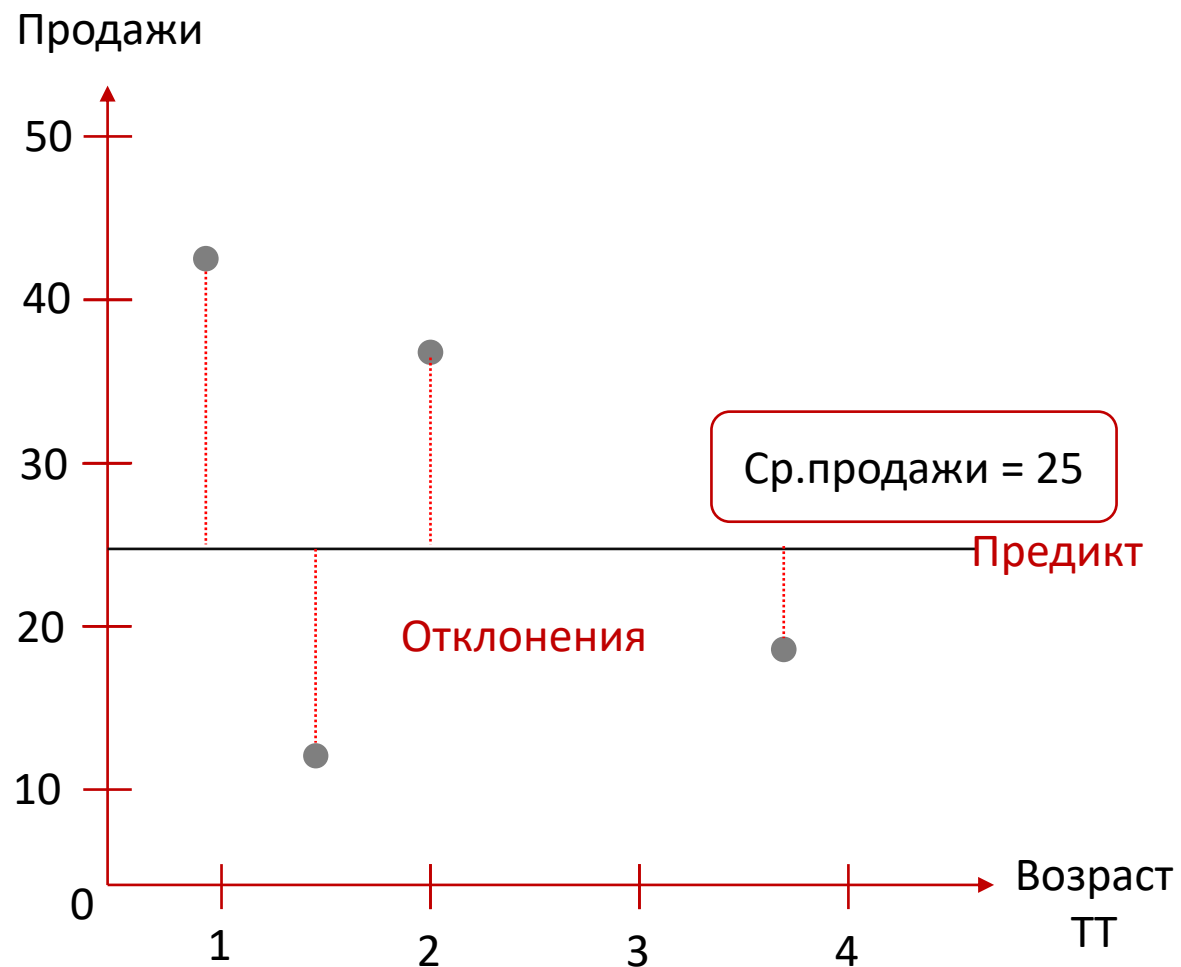
SIMILARITY SCORE



19, -13, 13, -6

$$\text{Скоринг} = \frac{169}{4 + 0}$$

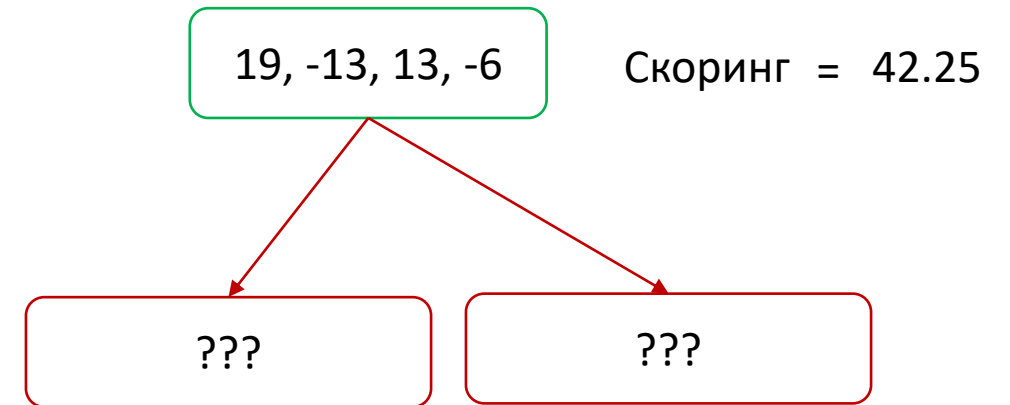
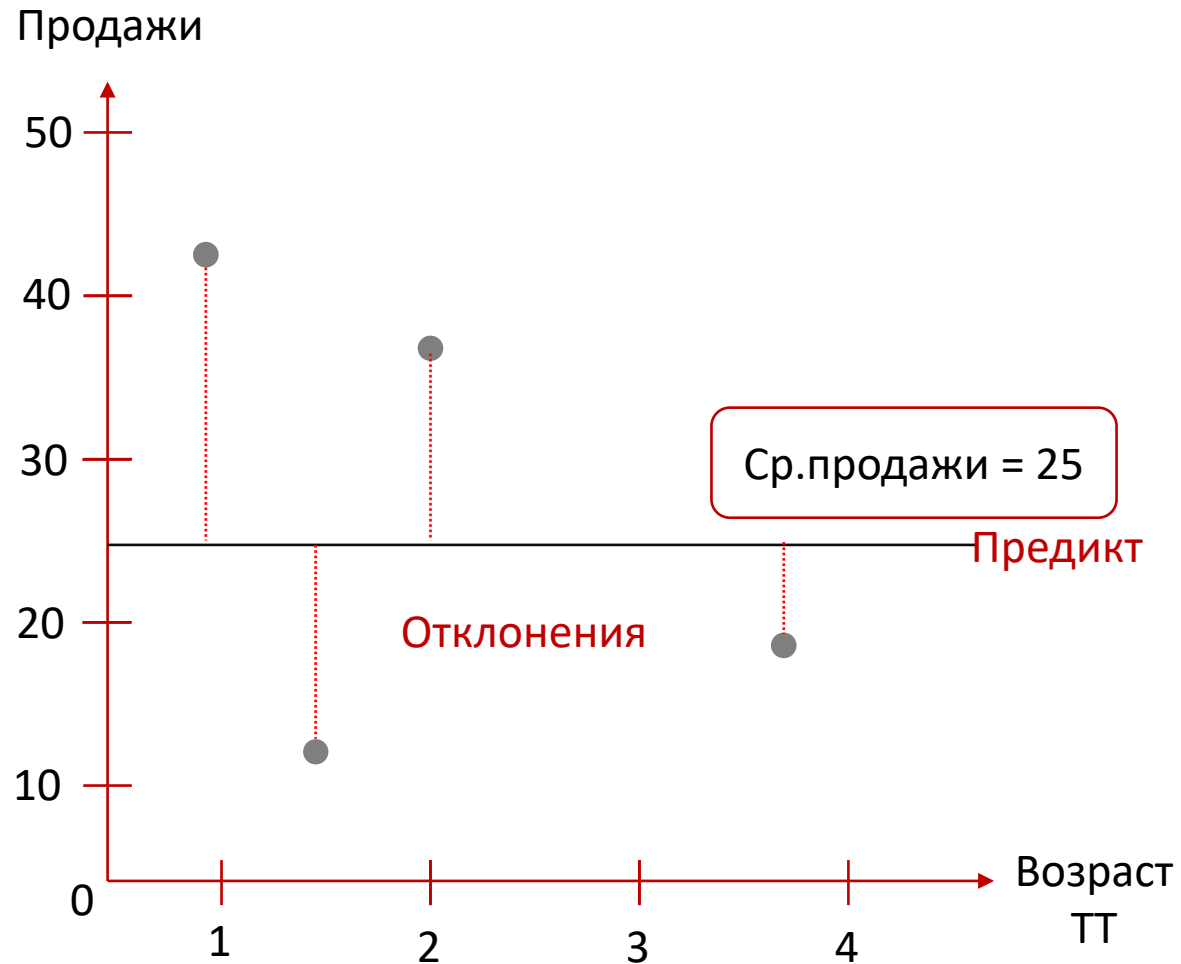
SIMILARITY SCORE



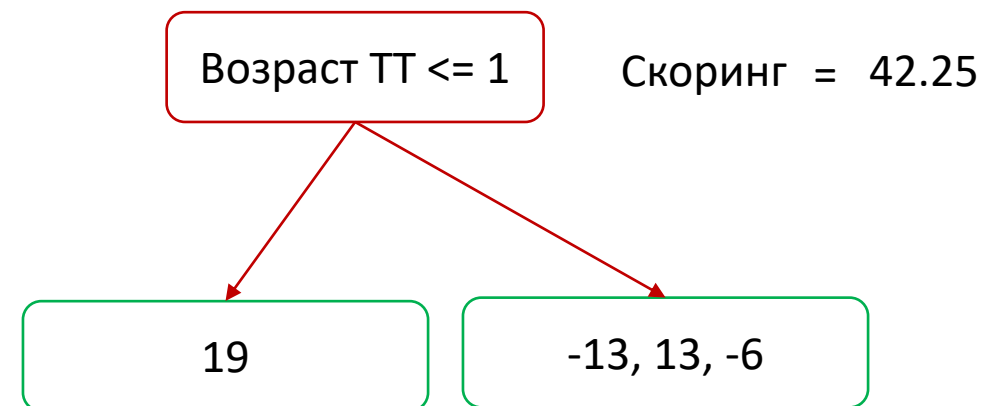
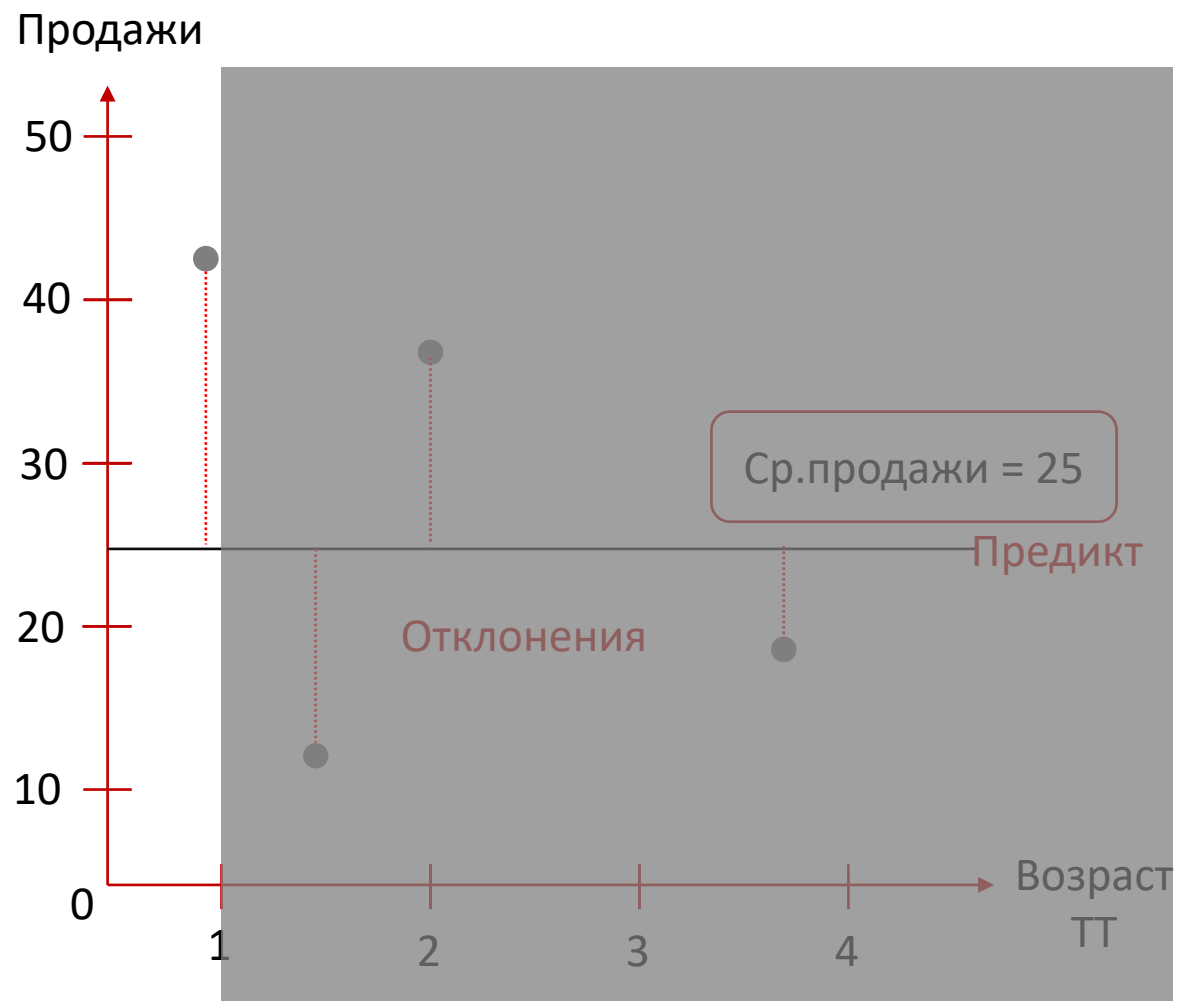
19, -13, 13, -6

Скоринг = 42.25

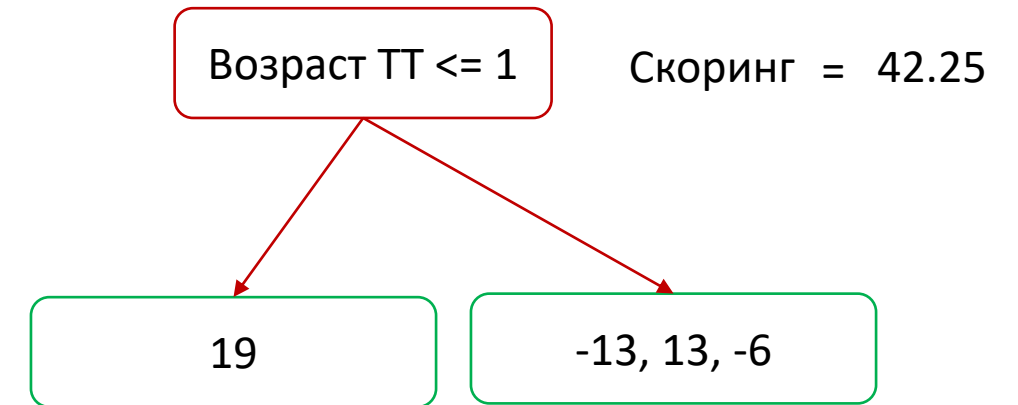
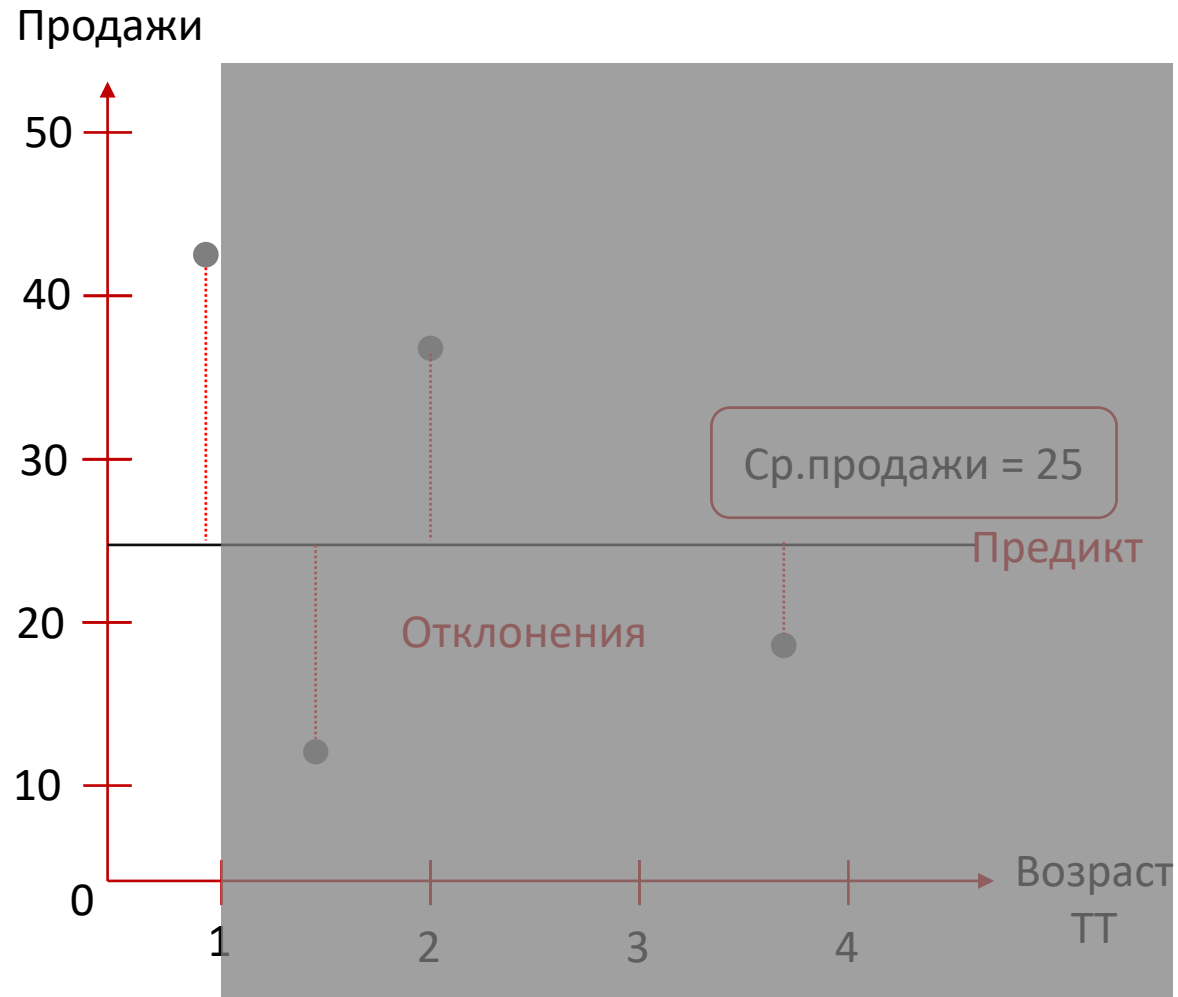
SIMILARITY SCORE



SIMILARITY SCORE

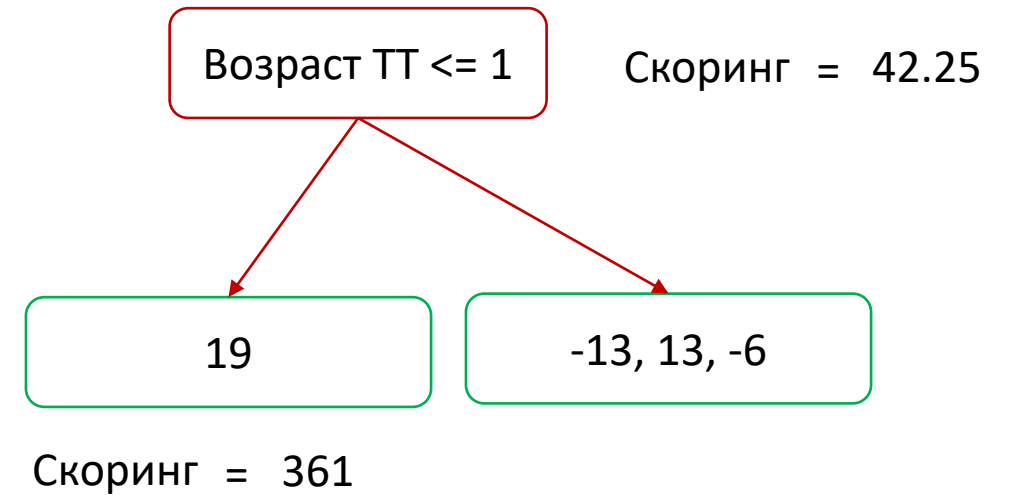
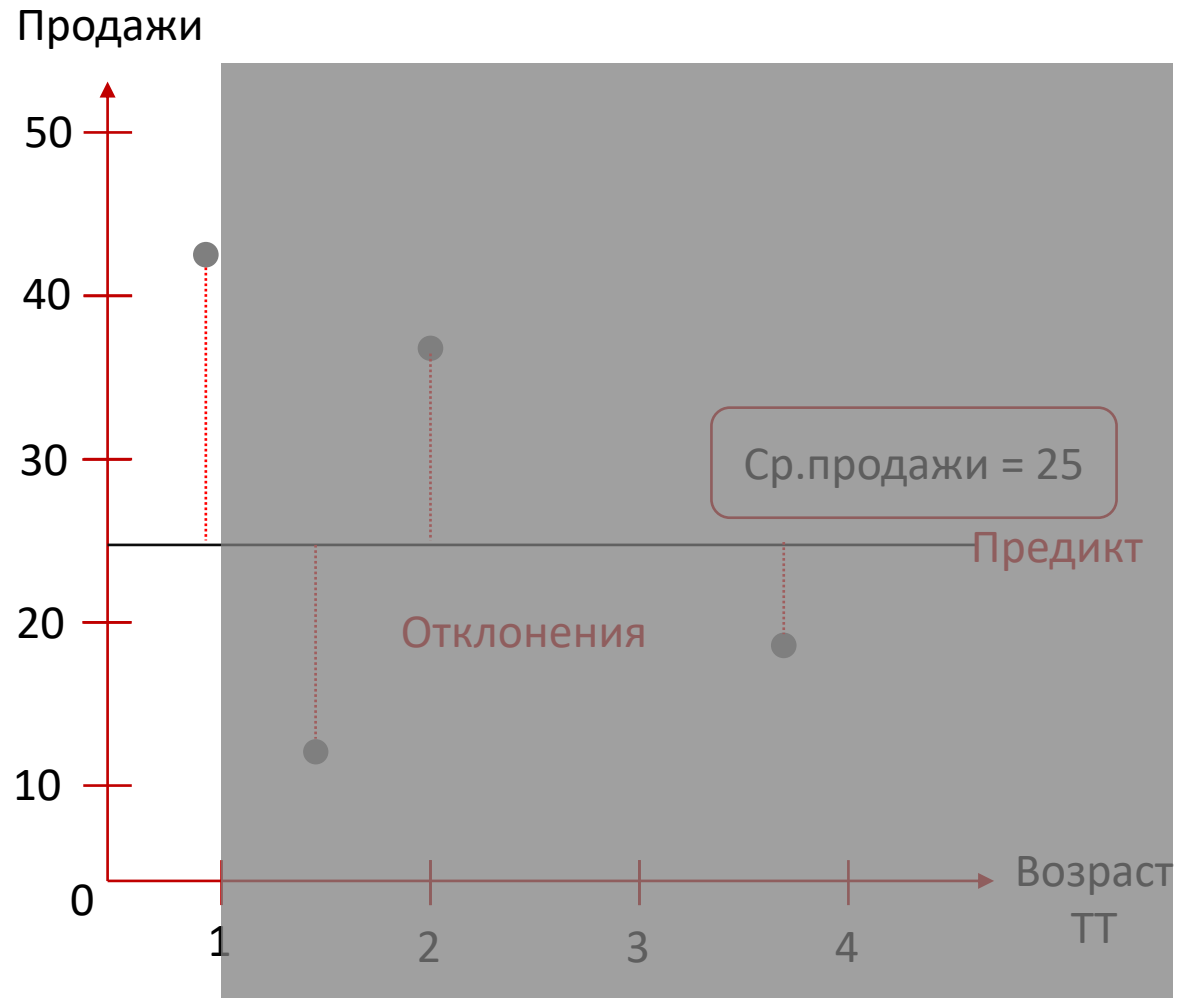


SIMILARITY SCORE

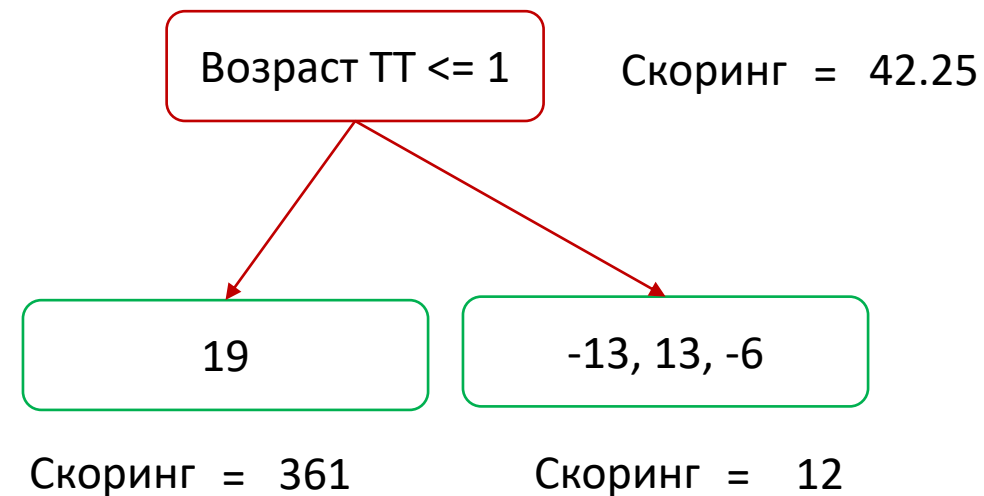
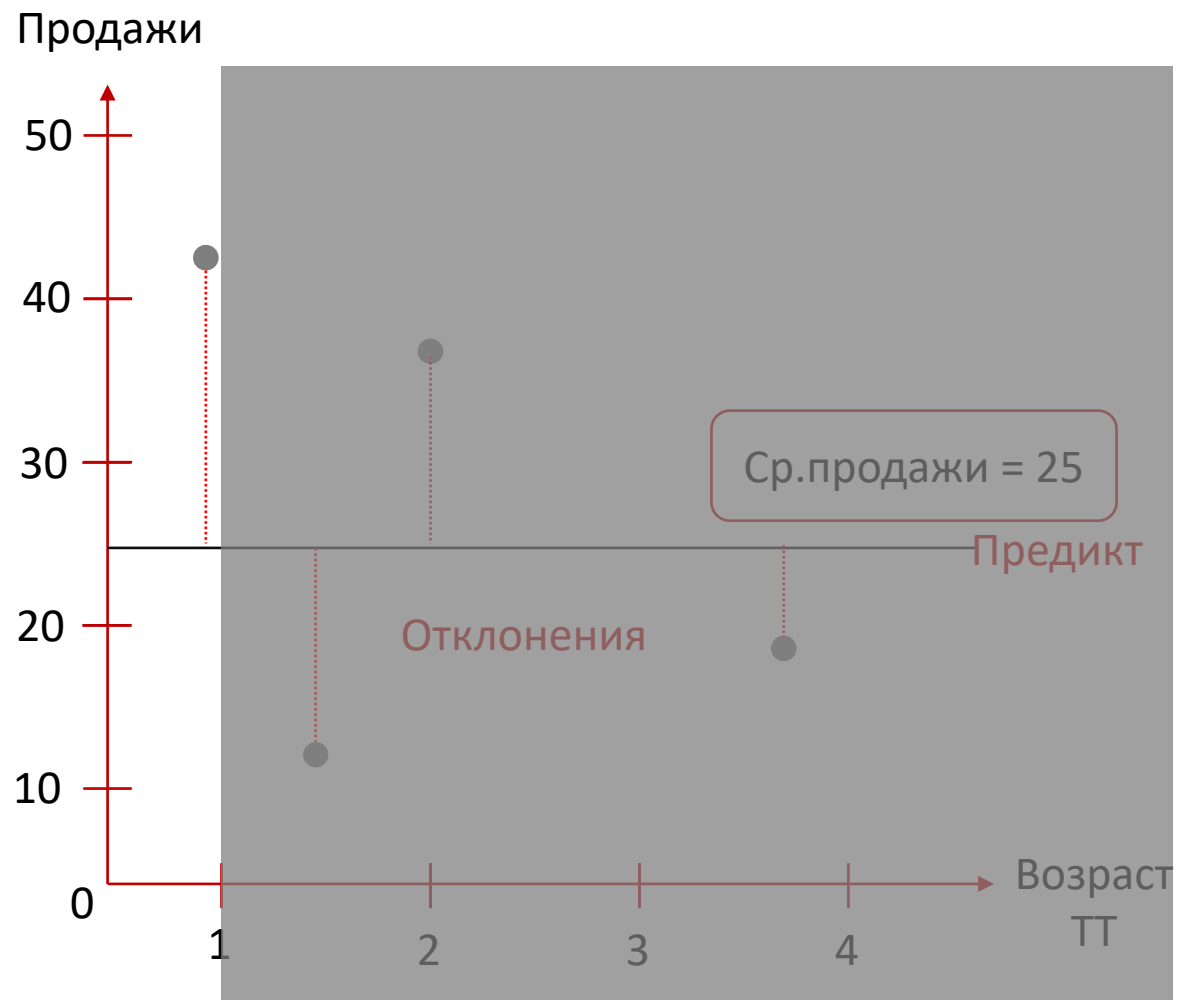


$$\text{Скоринг} = \frac{19^{**}2}{1 + 0}$$

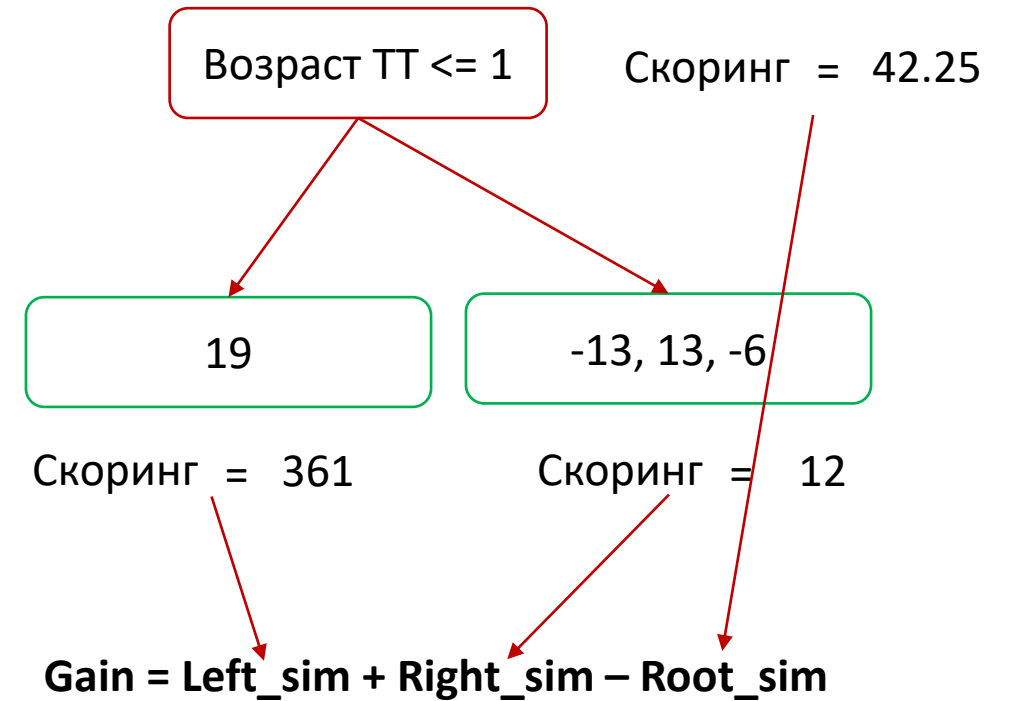
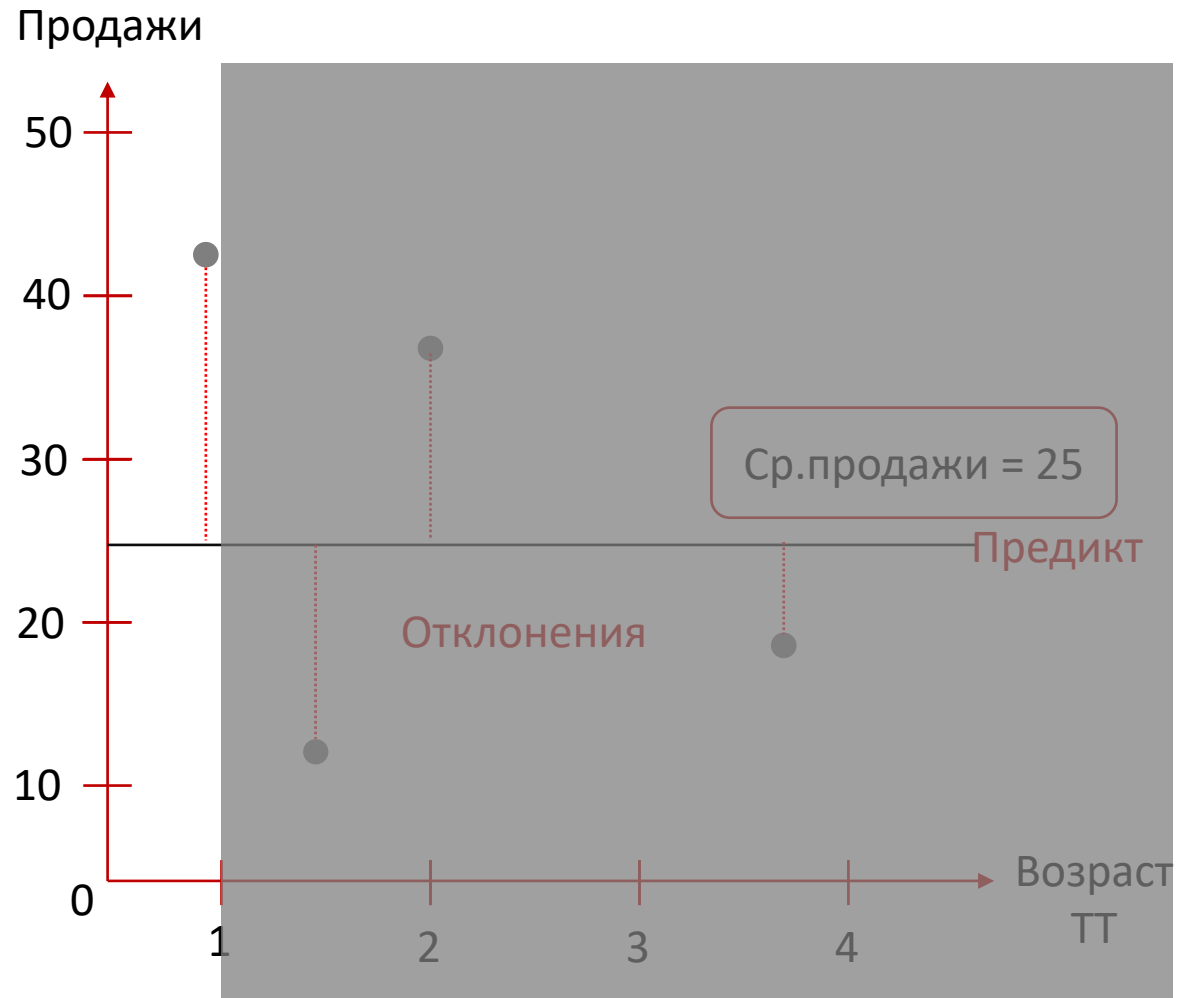
SIMILARITY SCORE



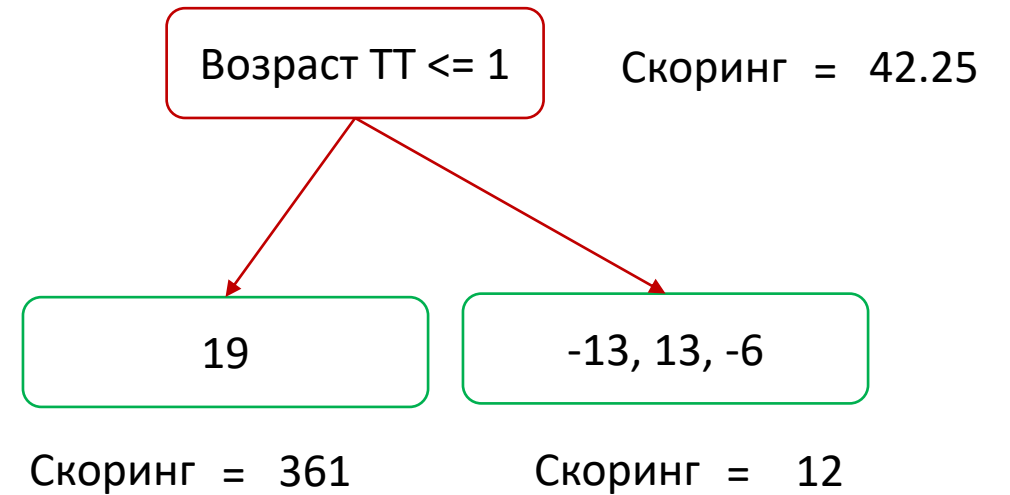
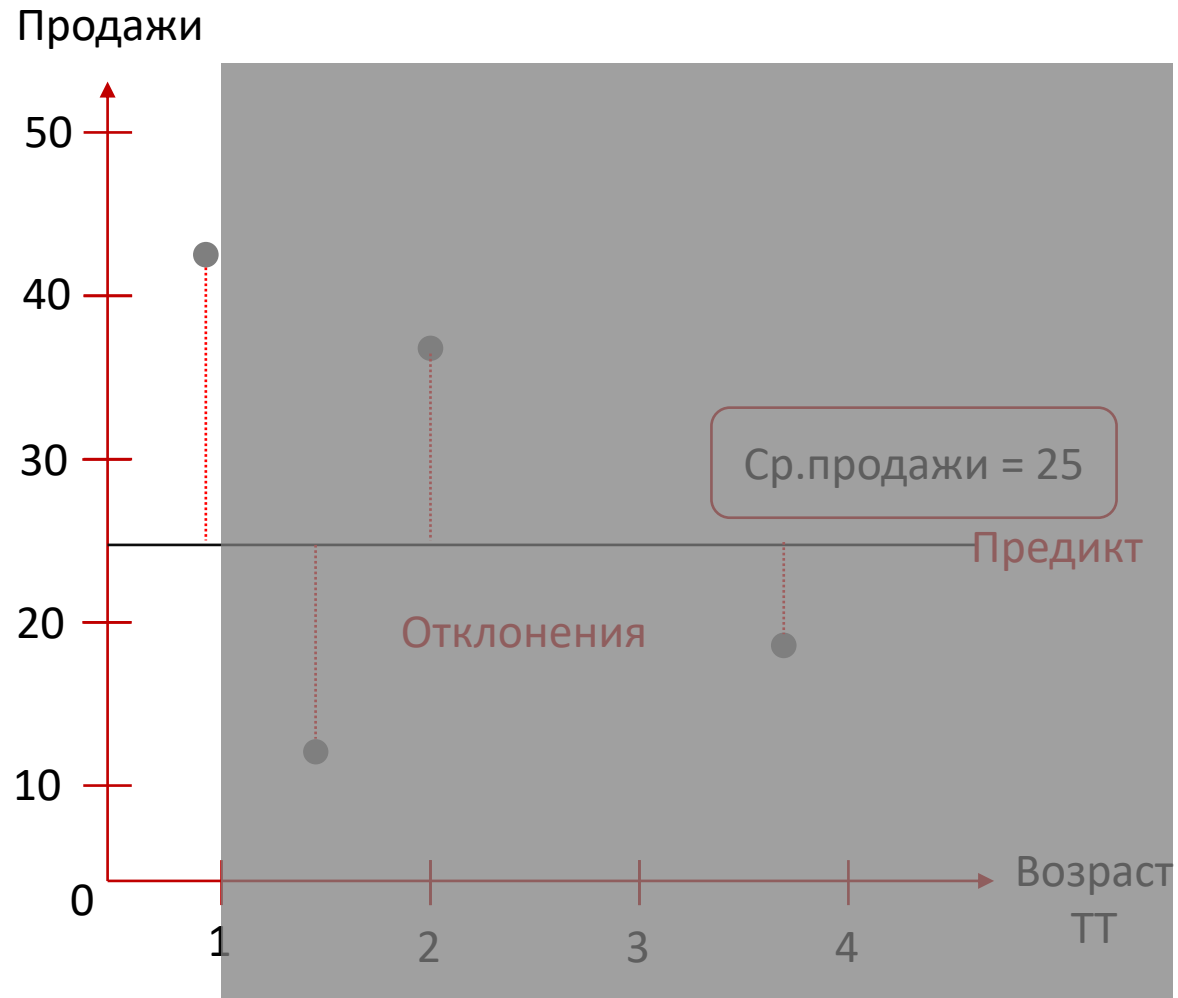
SIMILARITY SCORE



GAIN

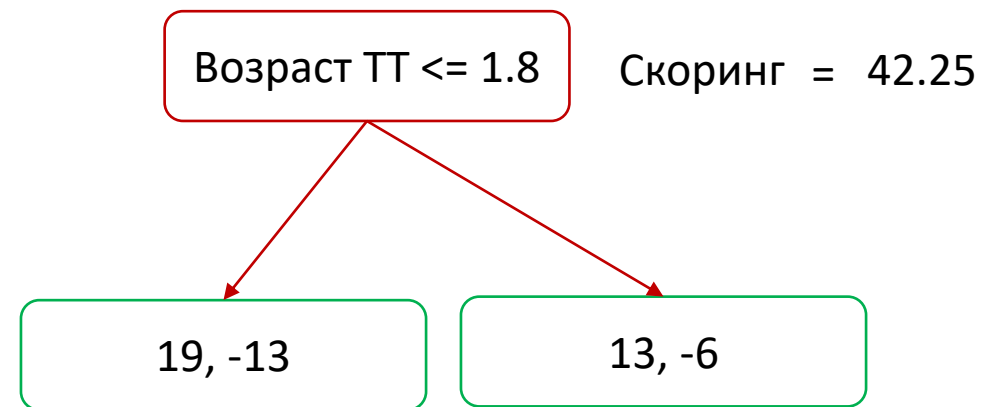
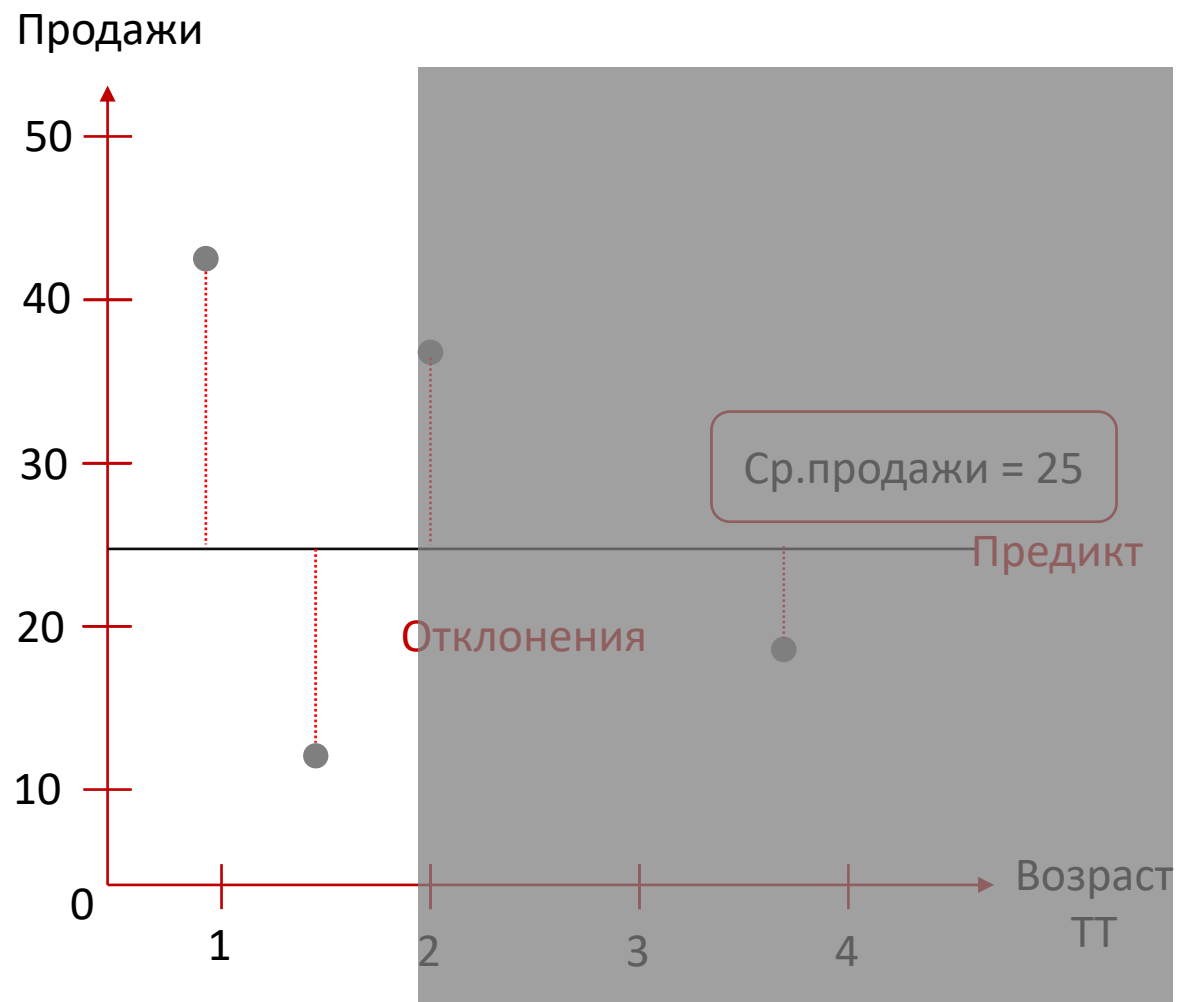


GAIN

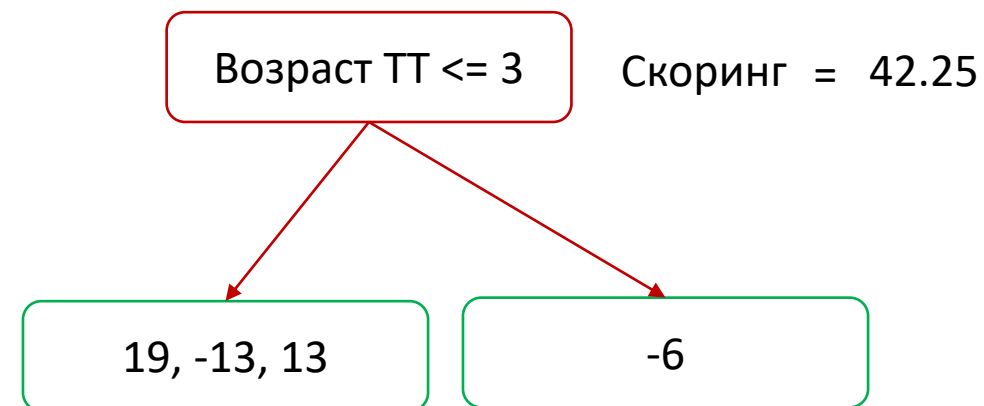
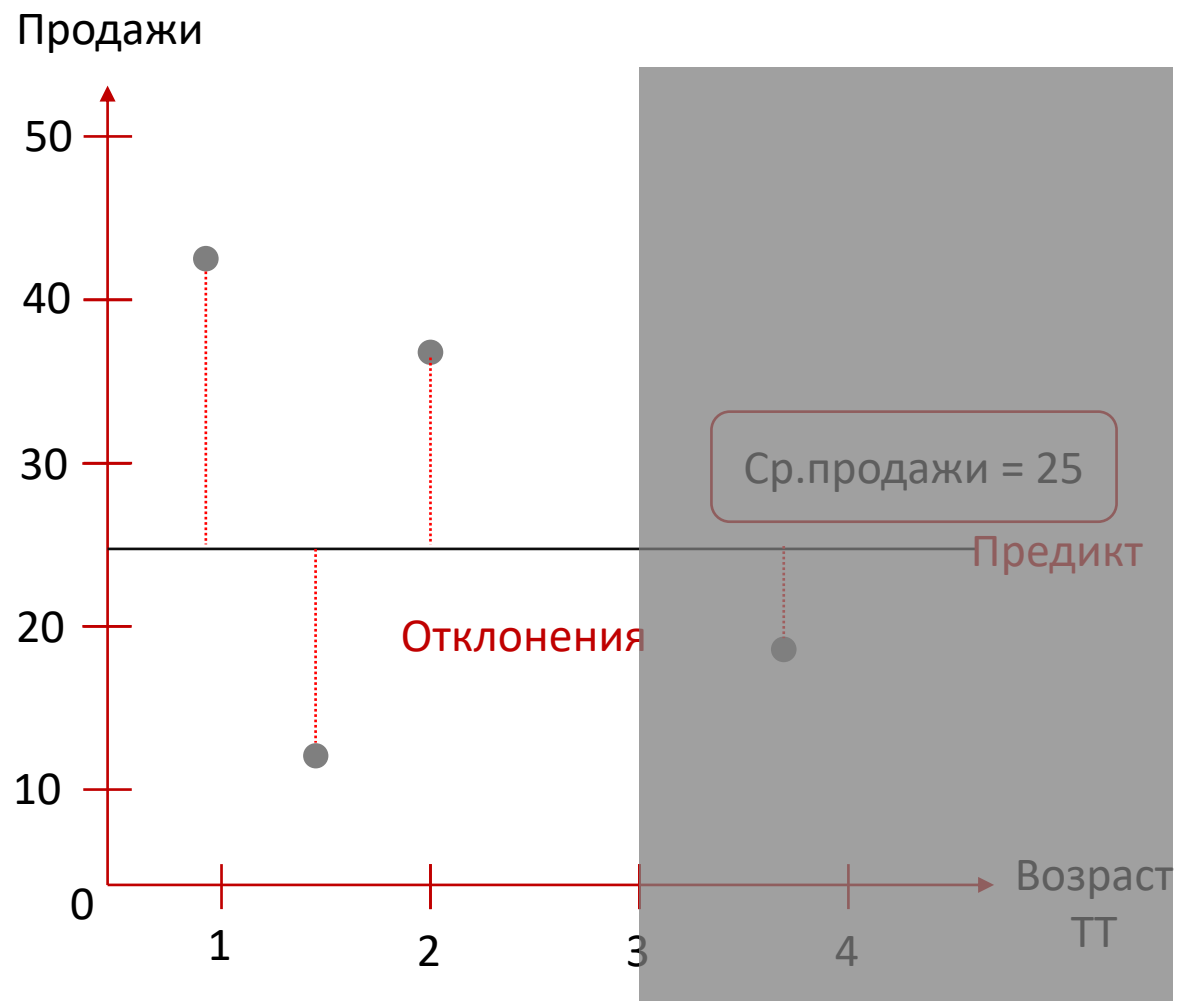


$$\text{Gain} = 361 + 12 - 42.25 = 330.75$$

GAIN

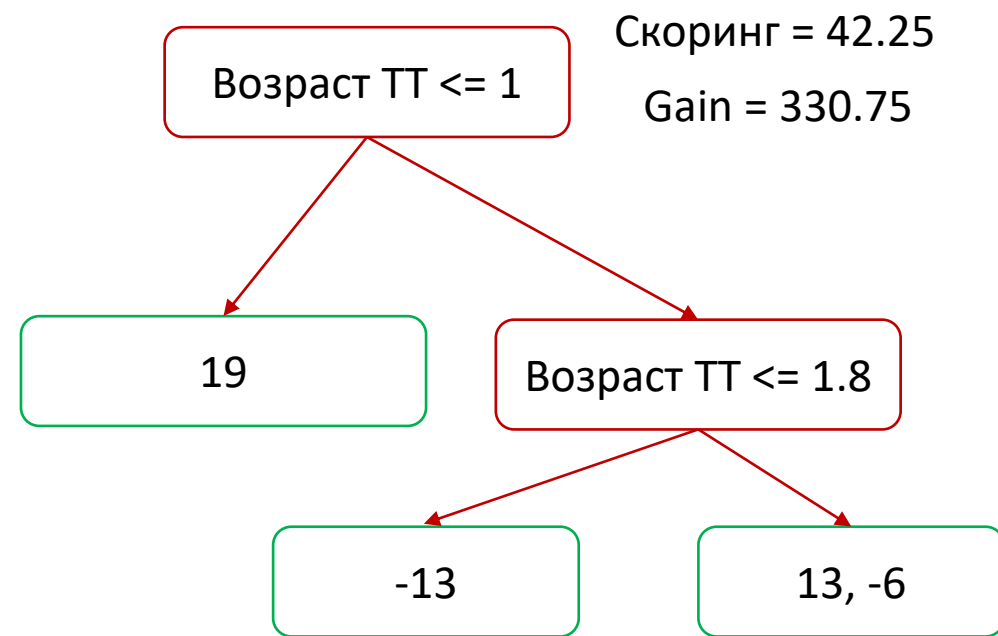
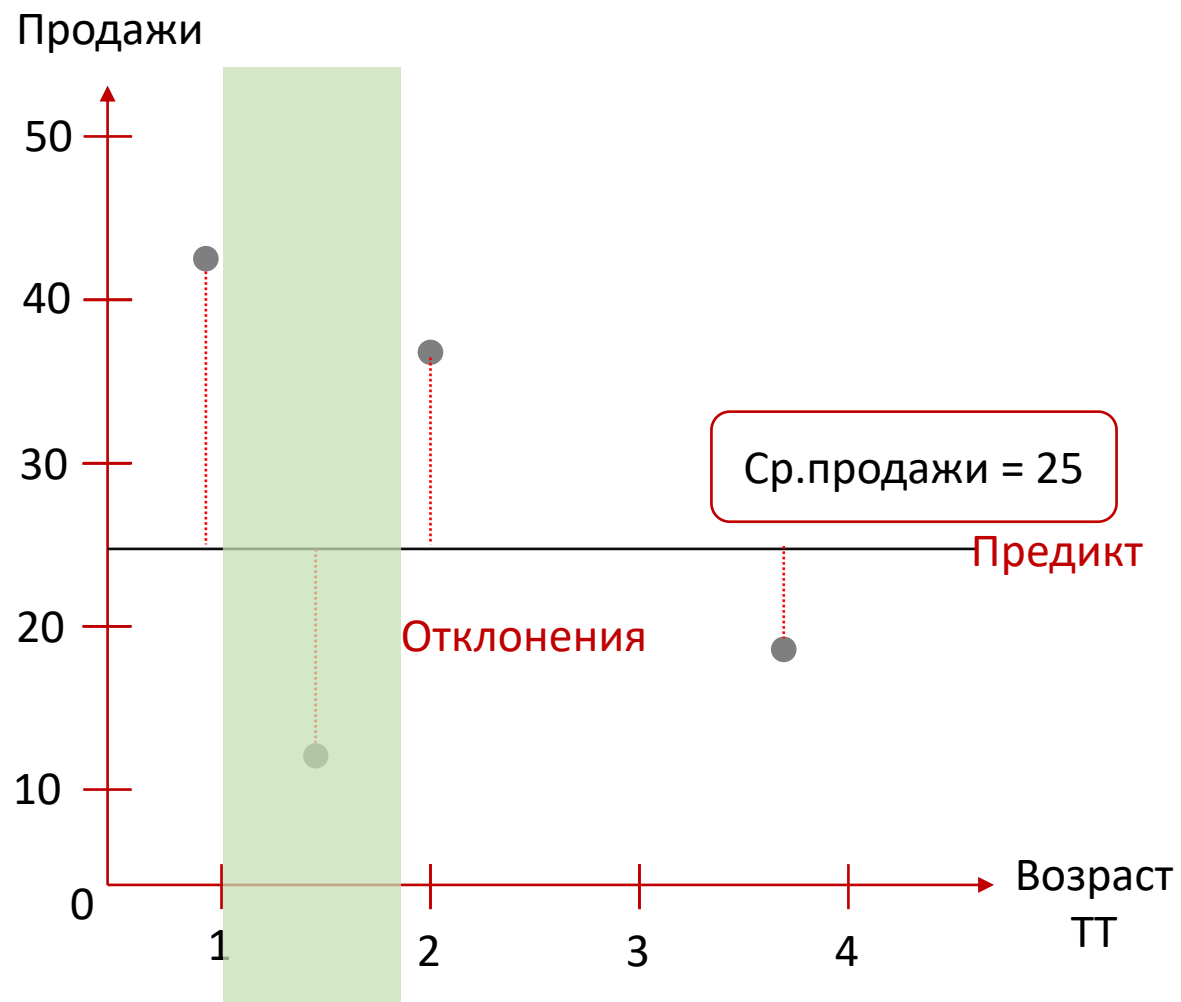


GAIN

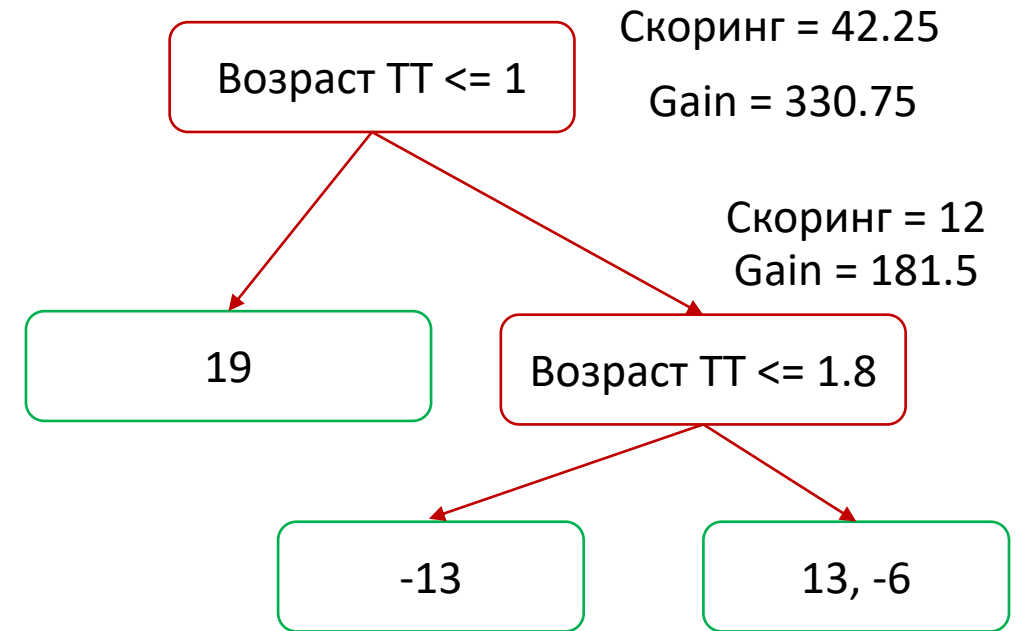
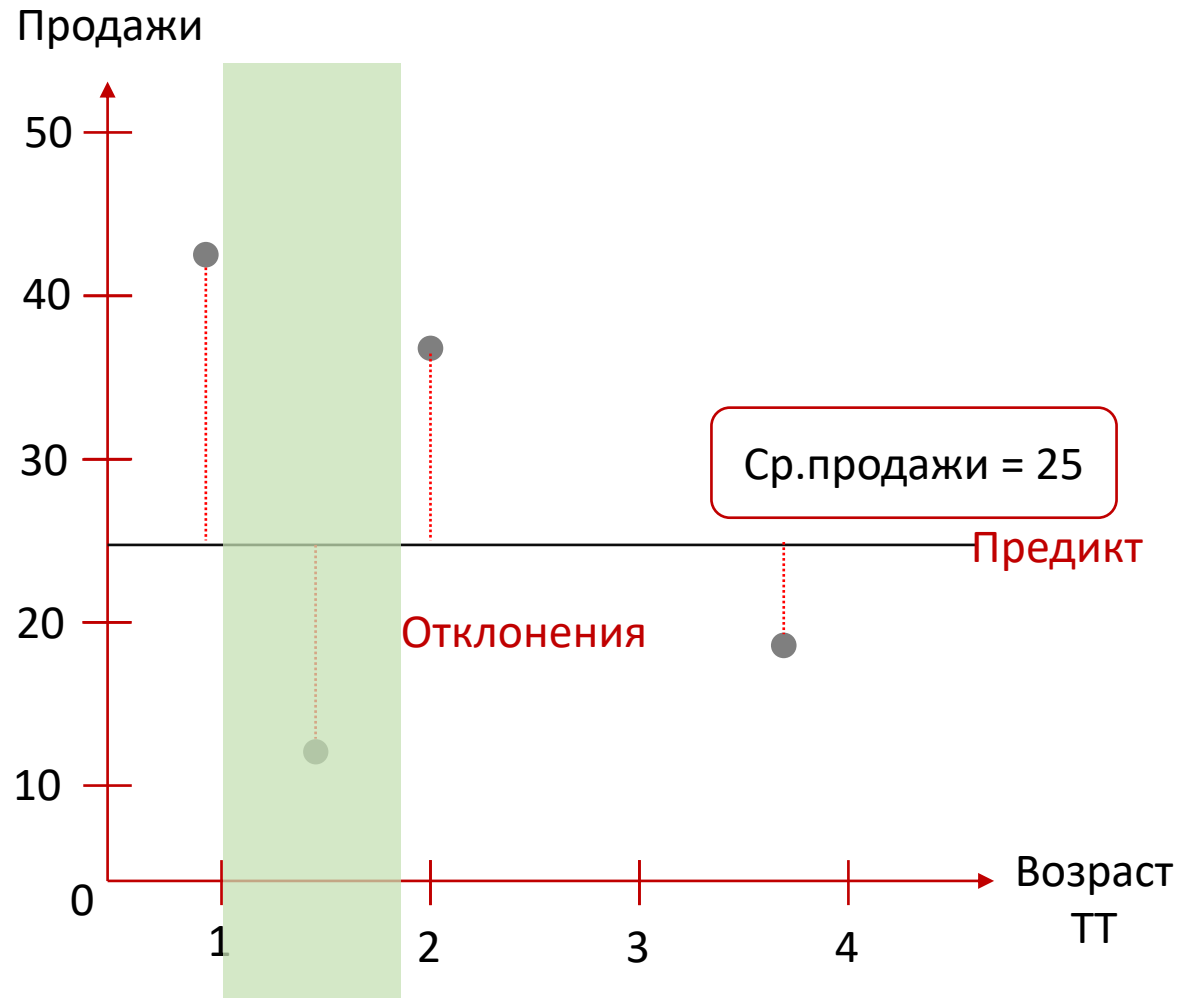


Находим наилучший GAIN и SCORE на каждой разбивке параметра

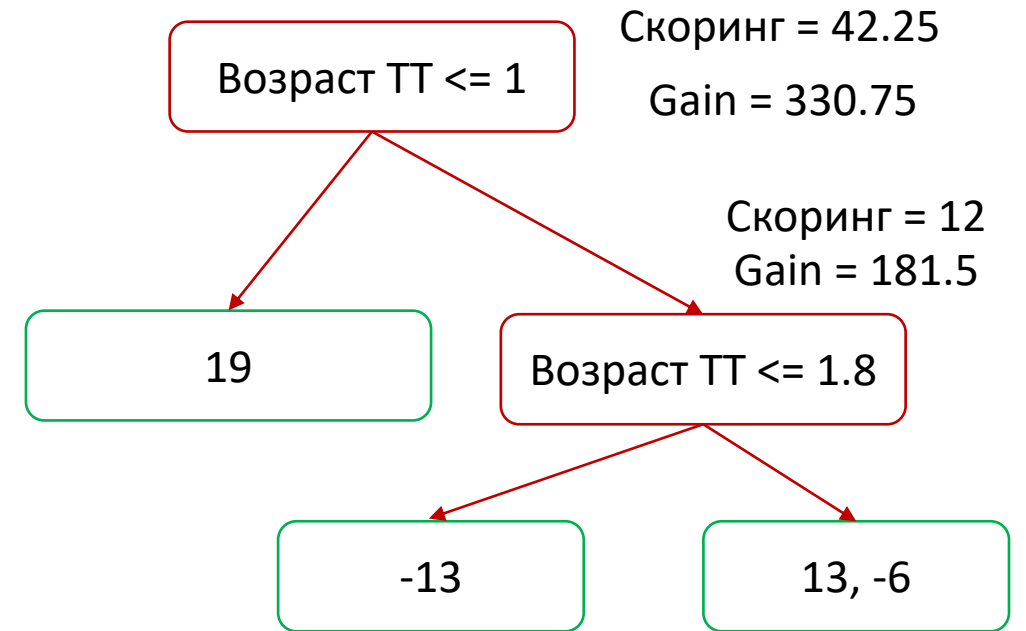
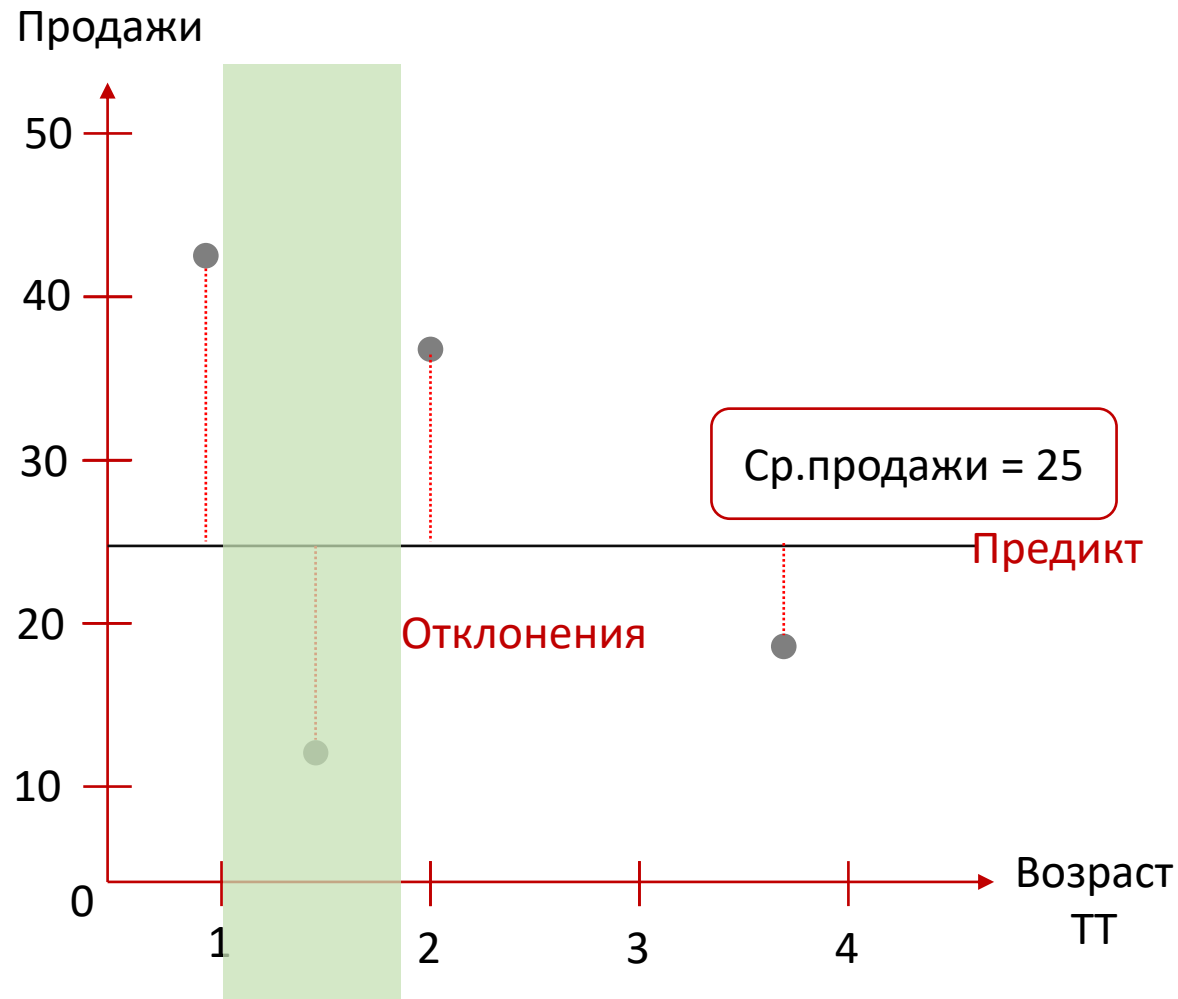
GAIN



PRUNE



PRUNE



Gamma для Pruning? Случайное число

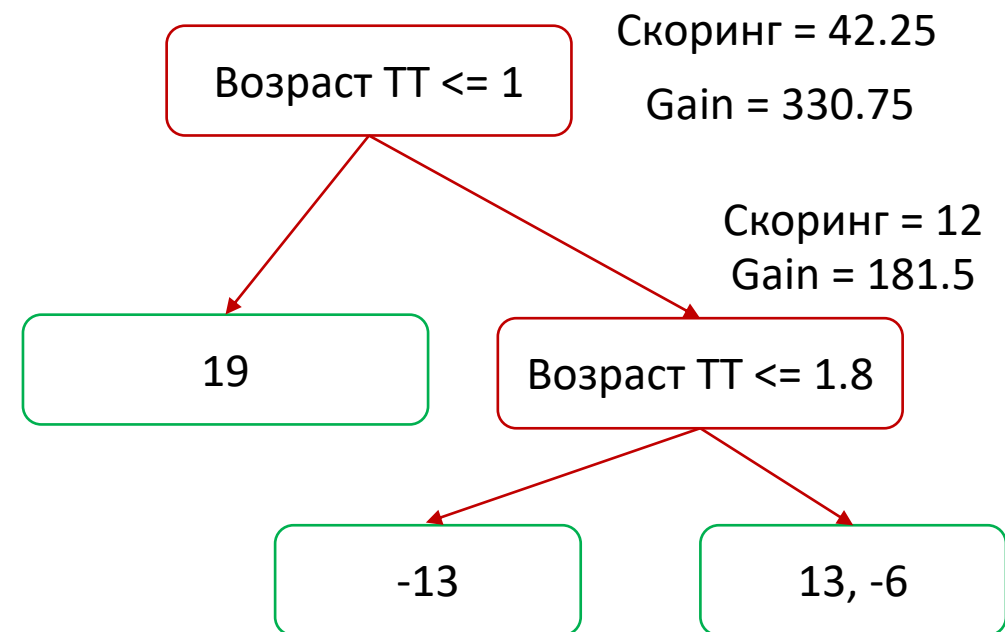
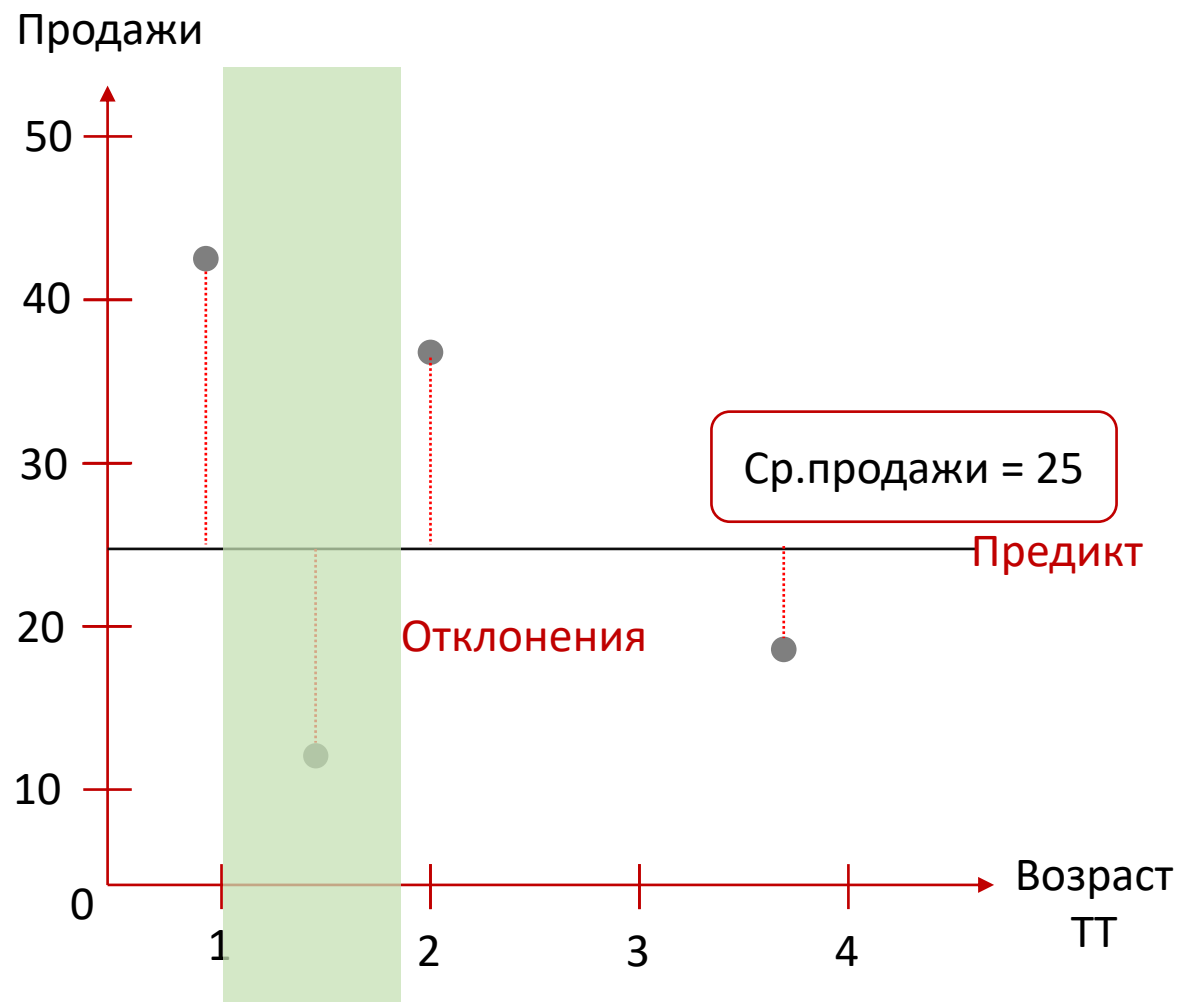
ЕСЛИ GAIN – Gamma < 0:

удалить

ИНАЧЕ:

считать дальше

PRUNE



Gamma для Pruning? Случайное число

1. $GAIN - Gamma = 330.75 - 200 = 130.75$
2. $GAIN - Gamma = 181.5 - 200 = -18.5$

STOP STOP STOP
А ПЕРВЫЙ ШАГ?



ПЕРВЫЙ ШАГ. НА ЧТО УЧИТЬСЯ?

$$Set(x_i, x_j)_{i=1}^n$$

Возраст ТТ	Продажи
1,00	44,00
1,50	12,00
2,00	38,00
3,80	19,00

Функция предикта

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$

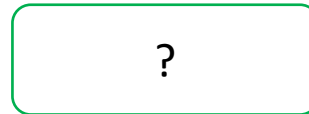
Функция потерь

$$\frac{1}{2}(\text{Observed} - \text{Predicted})^2 = \frac{2}{2}(\text{Observed} - \text{Predicted})x - 1 = -(\text{Observed} - \text{Predicted})$$

ПЕРВЫЙ ШАГ. НА ЧТО УЧИТЬСЯ?

$$Set(x_i, x_j)_{i=1}^n$$

Возраст ТТ	Продажи
1,00	44,00
1,50	12,00
2,00	38,00
3,80	19,00



ПЕРВЫЙ ШАГ. НА ЧТО УЧИТЬСЯ?

$$Set(x_i, x_j)_{i=1}^n$$

Возраст ТТ	Продажи
1,00	44,00
1,50	12,00
2,00	38,00
3,80	19,00

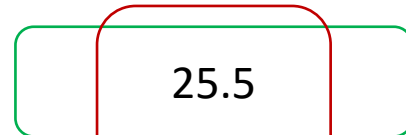
$$F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$$

AVG (ПРОДАЖИ)

ПЕРВЫЙ ШАГ. НА ЧТО УЧИТЬСЯ?

$Set(x_i, x_j)_{i=1}^n$

Возраст ТТ	Продажи
1,00	44,00
1,50	12,00
2,00	38,00
3,80	19,00



$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$

ПЕРВЫЙ ШАГ. НА ЧТО УЧИТЬСЯ?

$Set(x_i, x_j)_{i=1}^n$

Возраст ТТ	Продажи
1,00	44,00
1,50	12,00
2,00	38,00
3,80	19,00



25.5



Delta
18.5
-13.5
12.5
-6.5

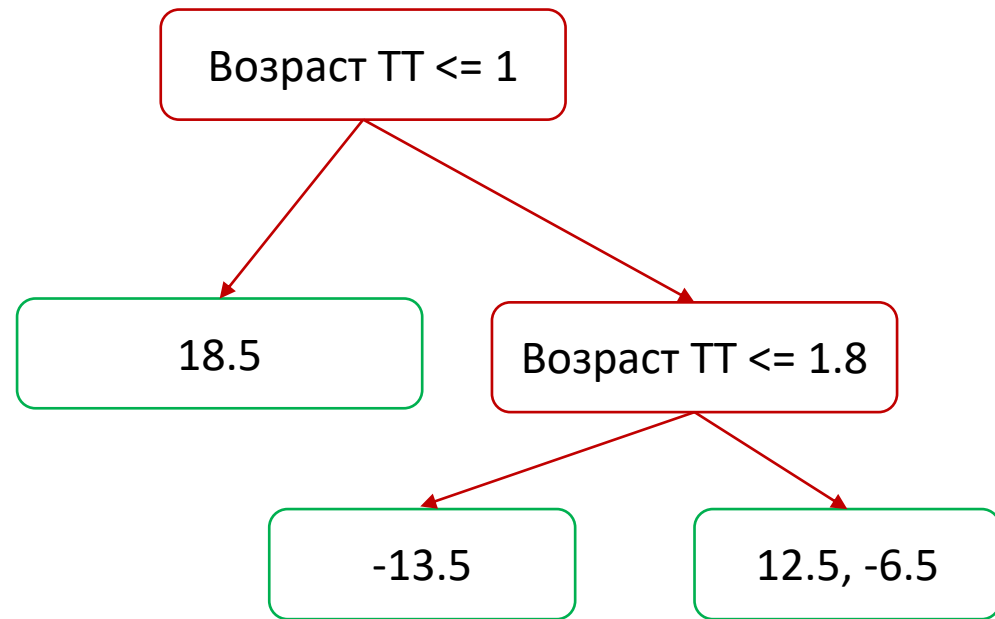
$$\frac{1}{2}(\text{Observed} - \text{Predicted})^2 = \frac{2}{2}(\text{Observed} - \text{Predicted})x - 1 = -(\text{Observed} - \text{Predicted})$$

ПЕРВЫЙ ШАГ. НА ЧТО УЧИТЬСЯ?

Возраст ТТ	Продажи	Delta
		18.5
1,00	44,00	-13.5
1,50	12,00	12.5
2,00	38,00	
3,80	19,00	-6.5

ПЕРВЫЙ ШАГ

Возраст ТТ	Продажи	Delta
1,00	44,00	-18.5
1,50	12,00	13.5
2,00	38,00	-12.5
3,80	19,00	6.5

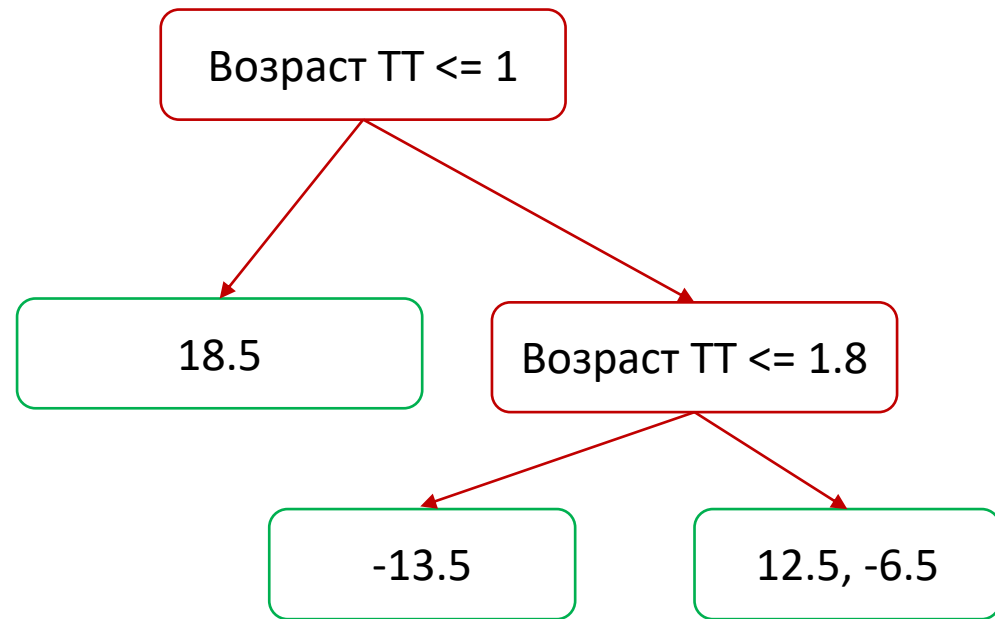


ПЕРВЫЙ ШАГ

Возраст ТТ	Продажи	Delta
1,00	44,00	-18.5
1,50	12,00	13.5
2,00	38,00	-12.5
3,80	19,00	6.5



Минимизируем отклонение
факт-предик

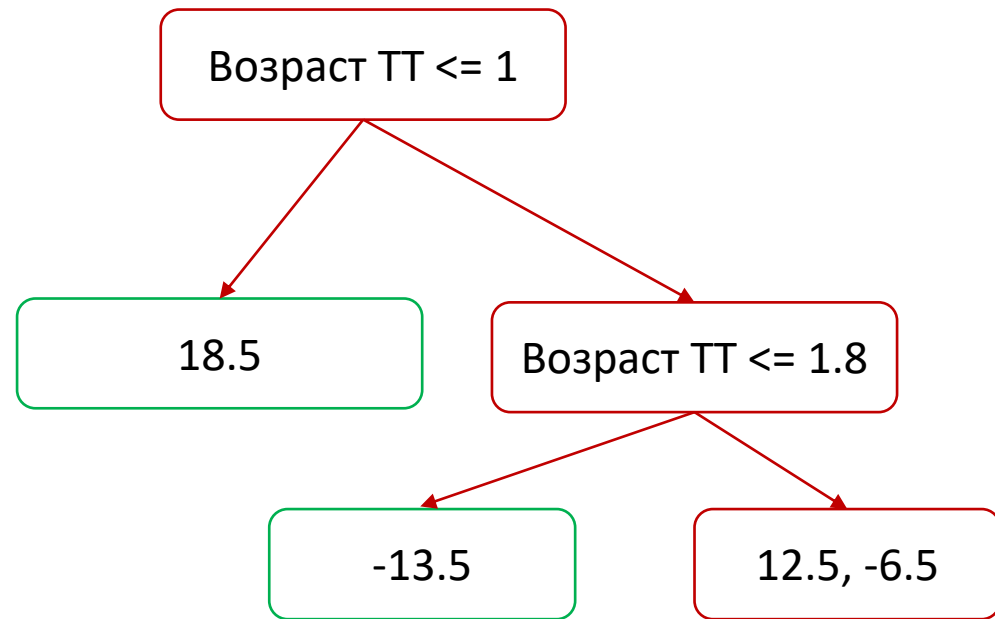


ПЕРВЫЙ ШАГ

Возраст ТТ	Продажи	Delta
		-18.5
1,00	44,00	13.5
1,50	12,00	-12.5
2,00	38,00	6.5
3,80	19,00	



Минимизируем отклонение
факт-предик

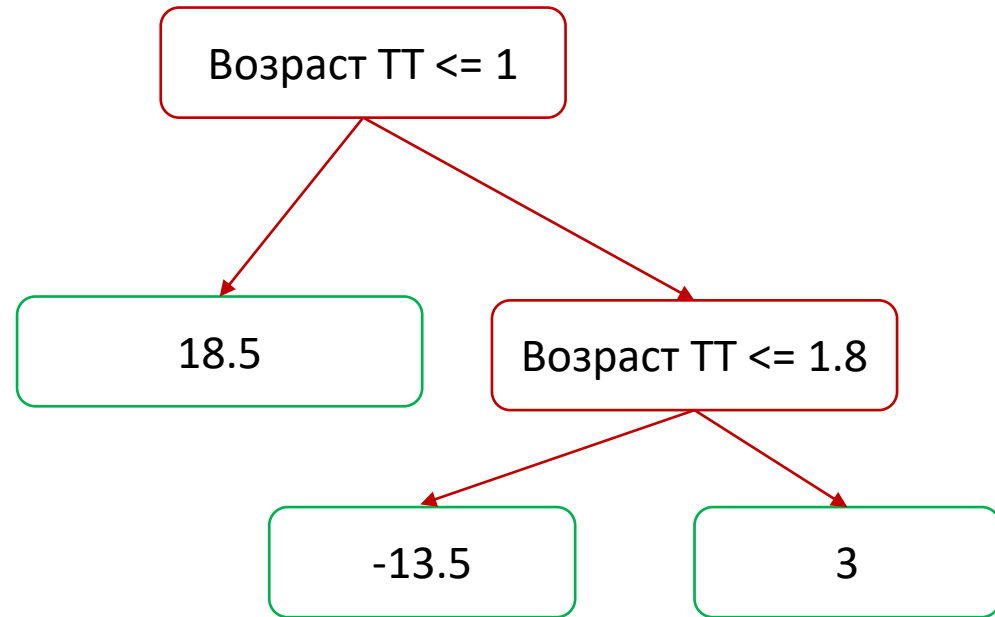


ПЕРВЫЙ ШАГ

Возраст ТТ	Продажи	Delta
1,00	44,00	-18.5
1,50	12,00	13.5
2,00	38,00	-12.5
3,80	19,00	6.5



Минимизируем отклонение
факт-предик



СПУСК

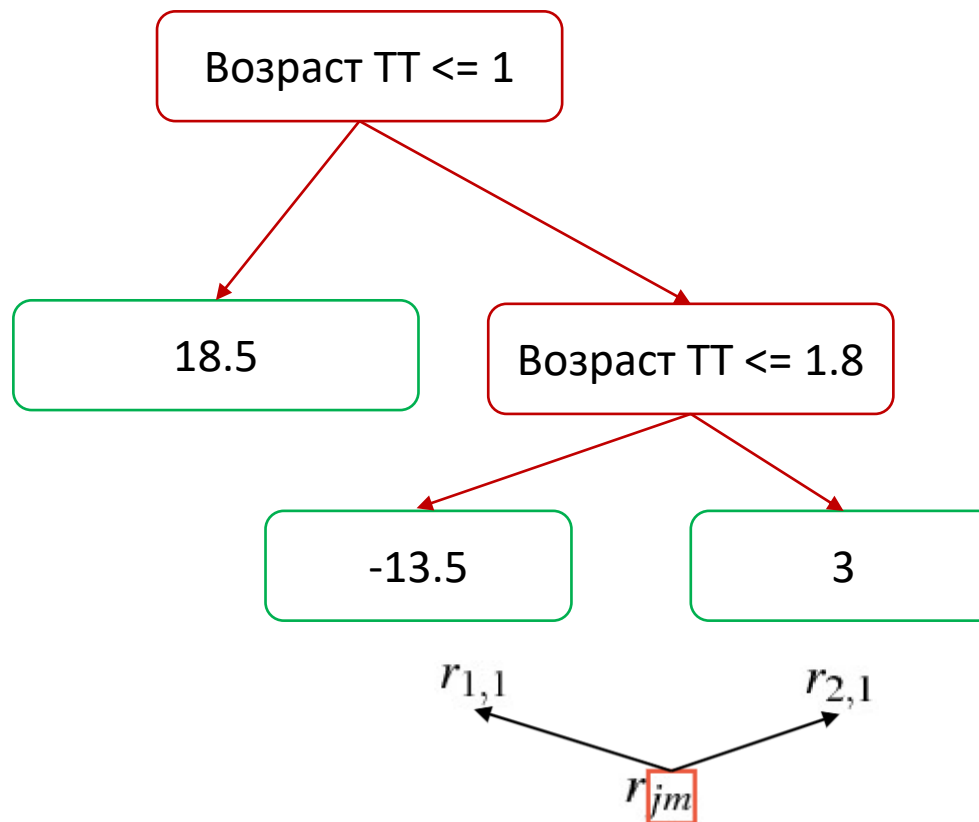


СПУСК

$$r_{i,m} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$

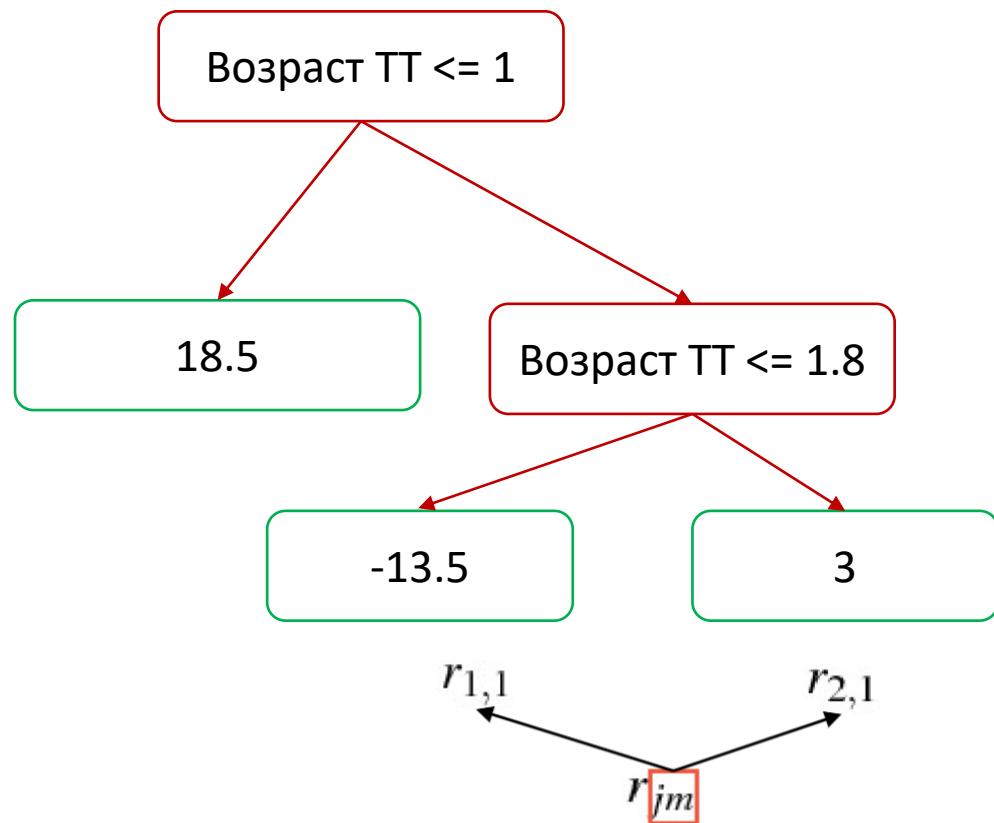
СПУСК

$$r_{i,m} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$



СПУСК

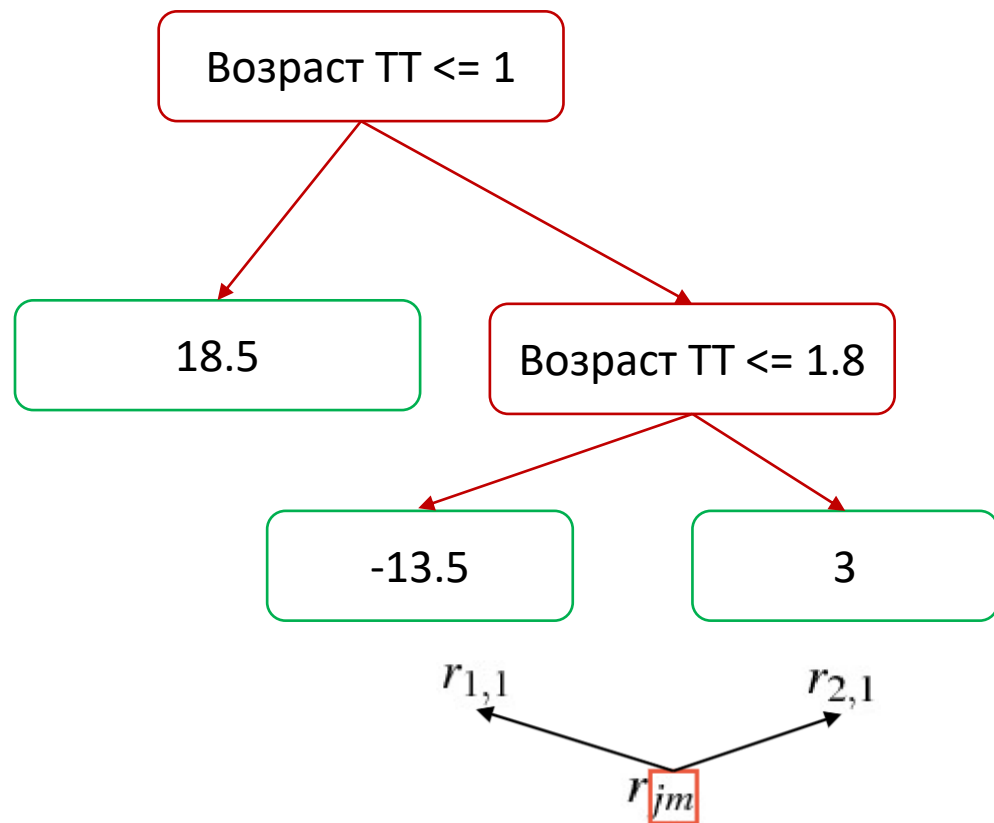
$$r_{i,m} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$



$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

СПУСК

$$r_{i,m} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$

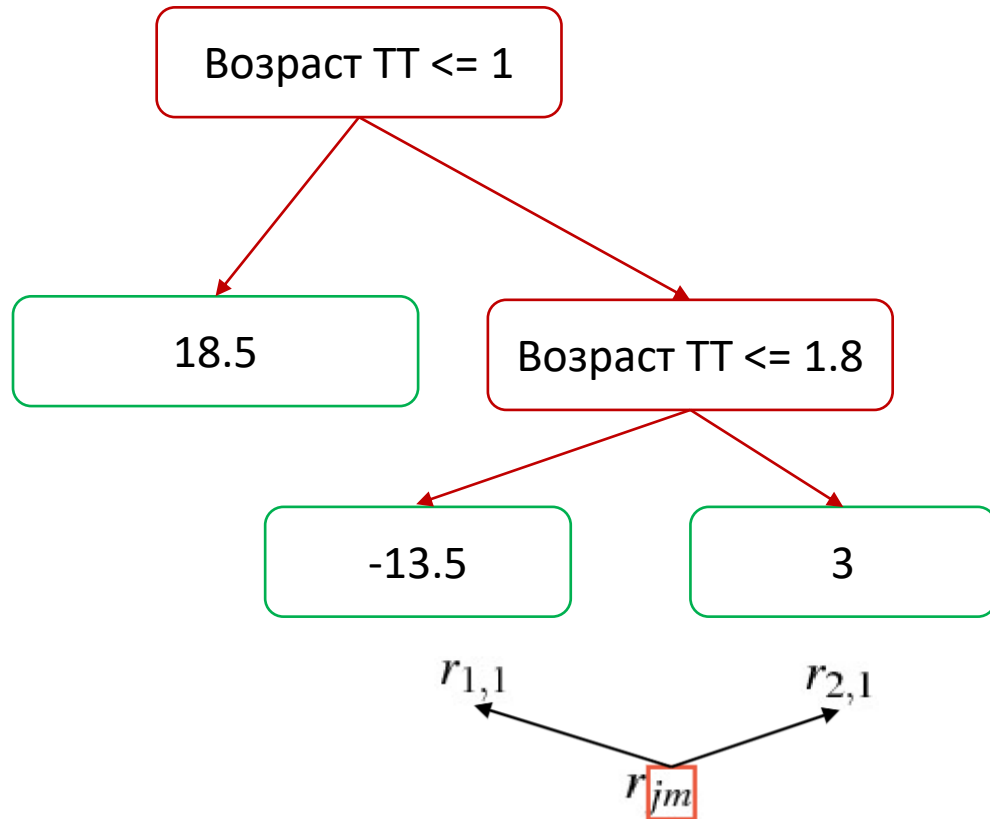


$$\gamma_{jm} = \text{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$F_0(x) = \text{argmin} \Sigma L(y_i, \gamma)$$

СПУСК

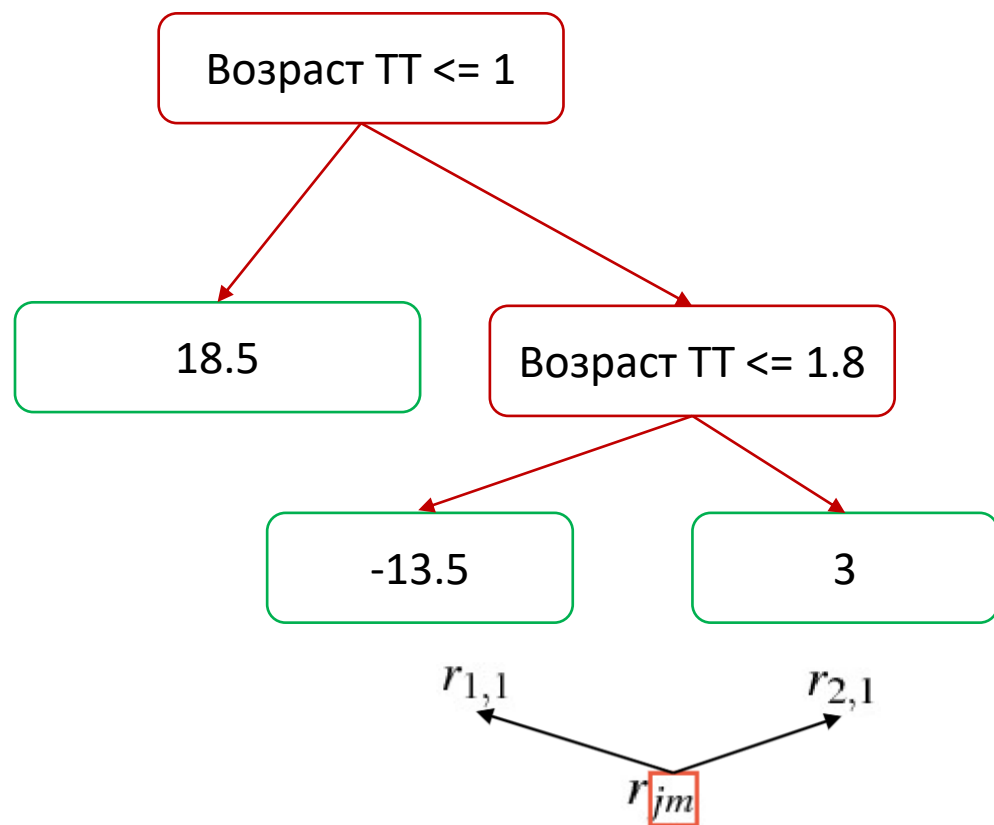
$$r_{i,m} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$



$$\text{Gamma}_{1,1} = \text{argmin} * 0.5 * (18.5 - \text{gamma})^{**2}$$

СПУСК

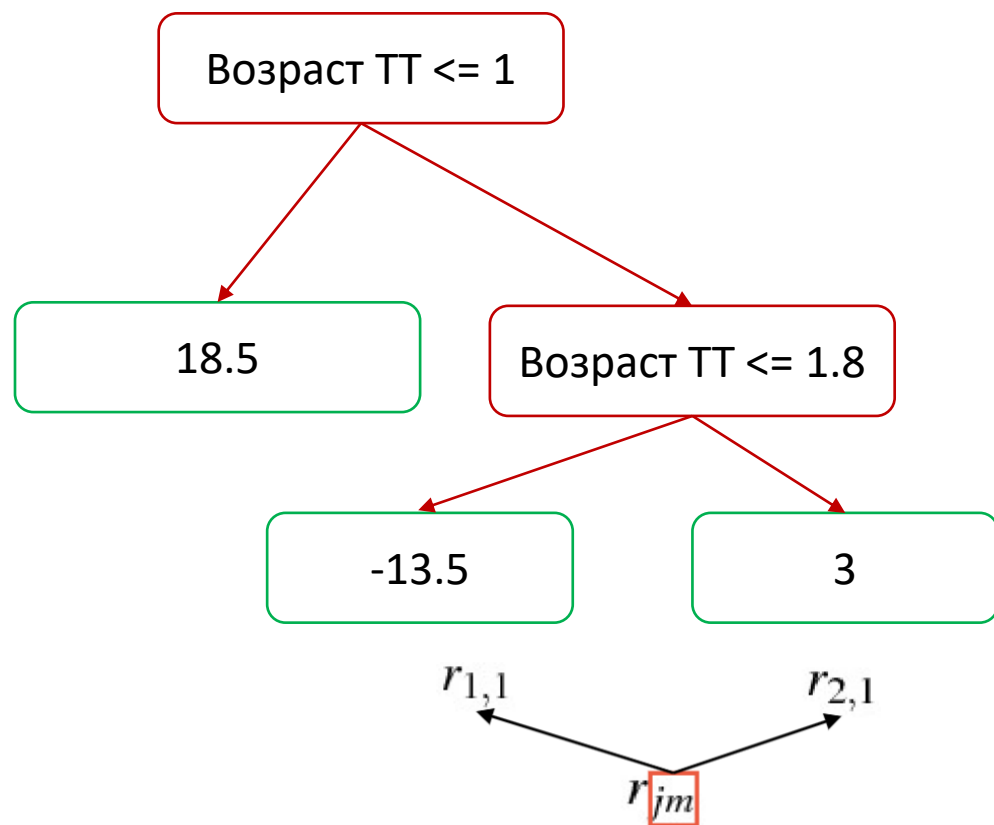
$$r_{i,m} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$



(18.5 - gamma)

СПУСК

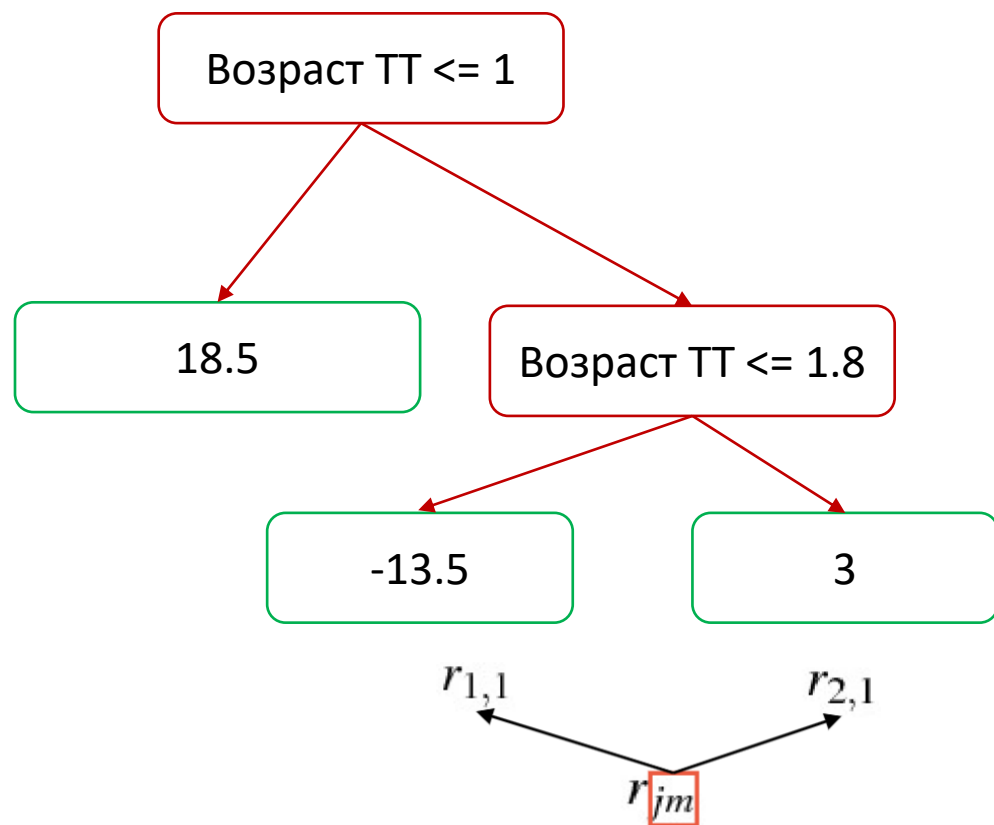
$$r_{i,m} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$



$$(\boxed{18.5} - 0)$$

СПУСК

$$r_{i,m} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$

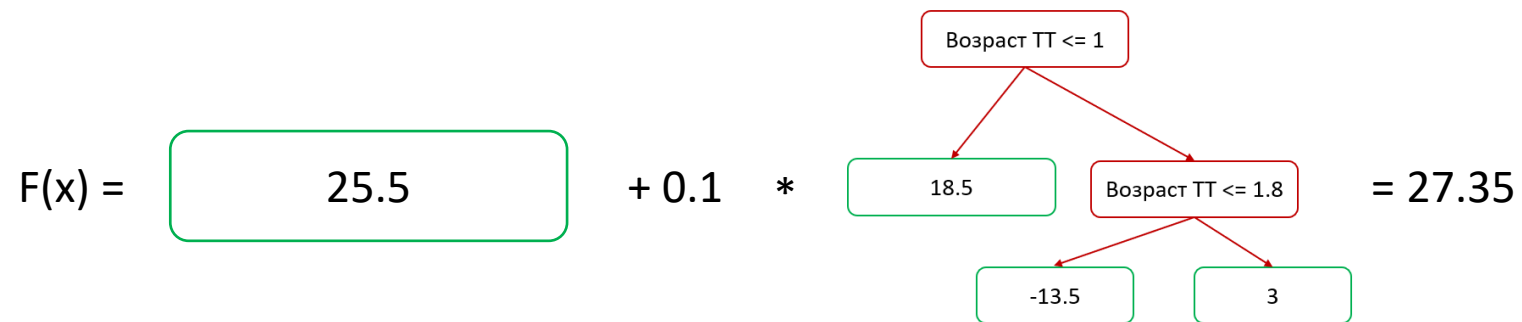


R_{i,m}=

18.5

ФИНАЛ

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$



ПЕРВЫЙ ПРЕДИКТ

Возраст ТТ	Продажи	Delta	Predict
		18.5	27.35
1,00	44,00	-13.5	26.85
1,50	12,00	12.5	26.1
2,00	38,00	-6.5	26.1
3,80	19,00		

ЗАДАНИЕ



entropy

Кому: Никита Симаков, Акимова Лилия

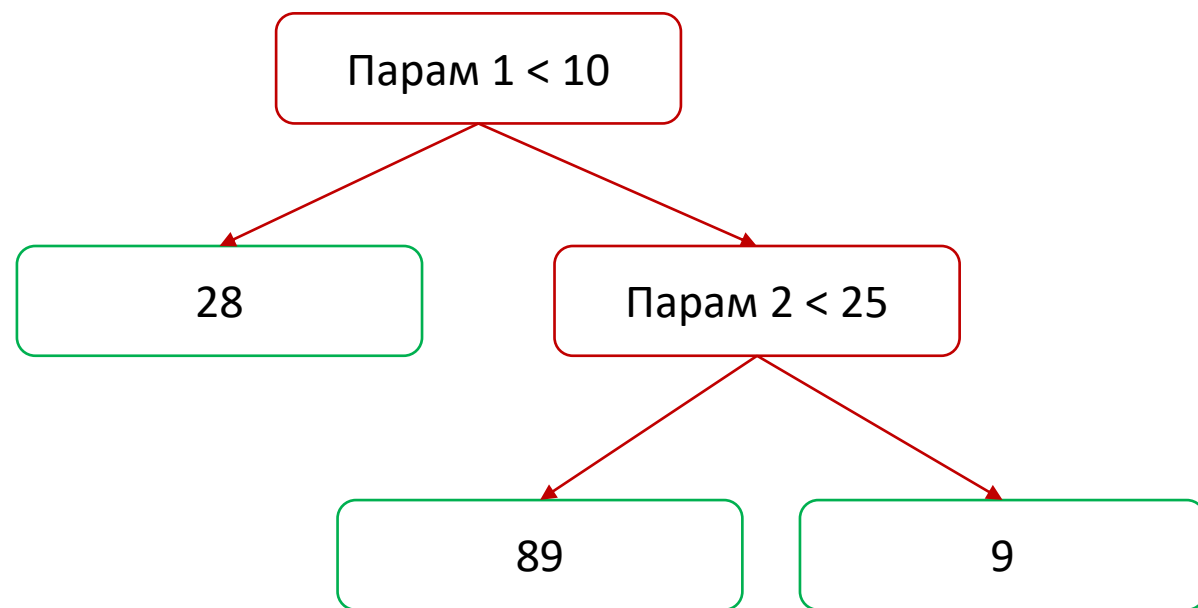
Презентация о Энтропии (с примером и python кодом)

BOOST В КЛАССИФИКАЦИИ



ПЛАН

1. Вспомним что было на Регрессии



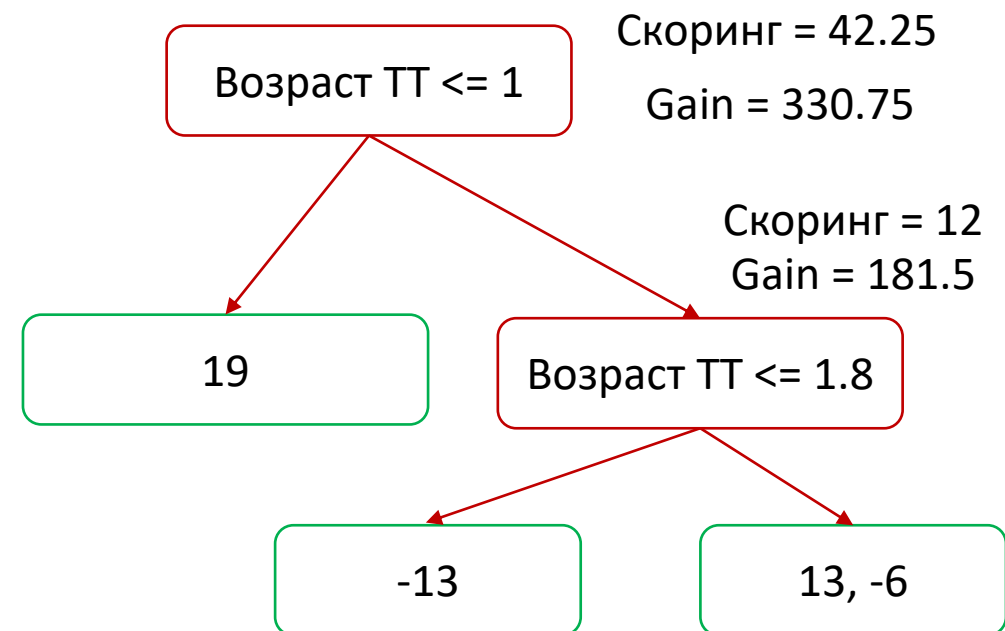
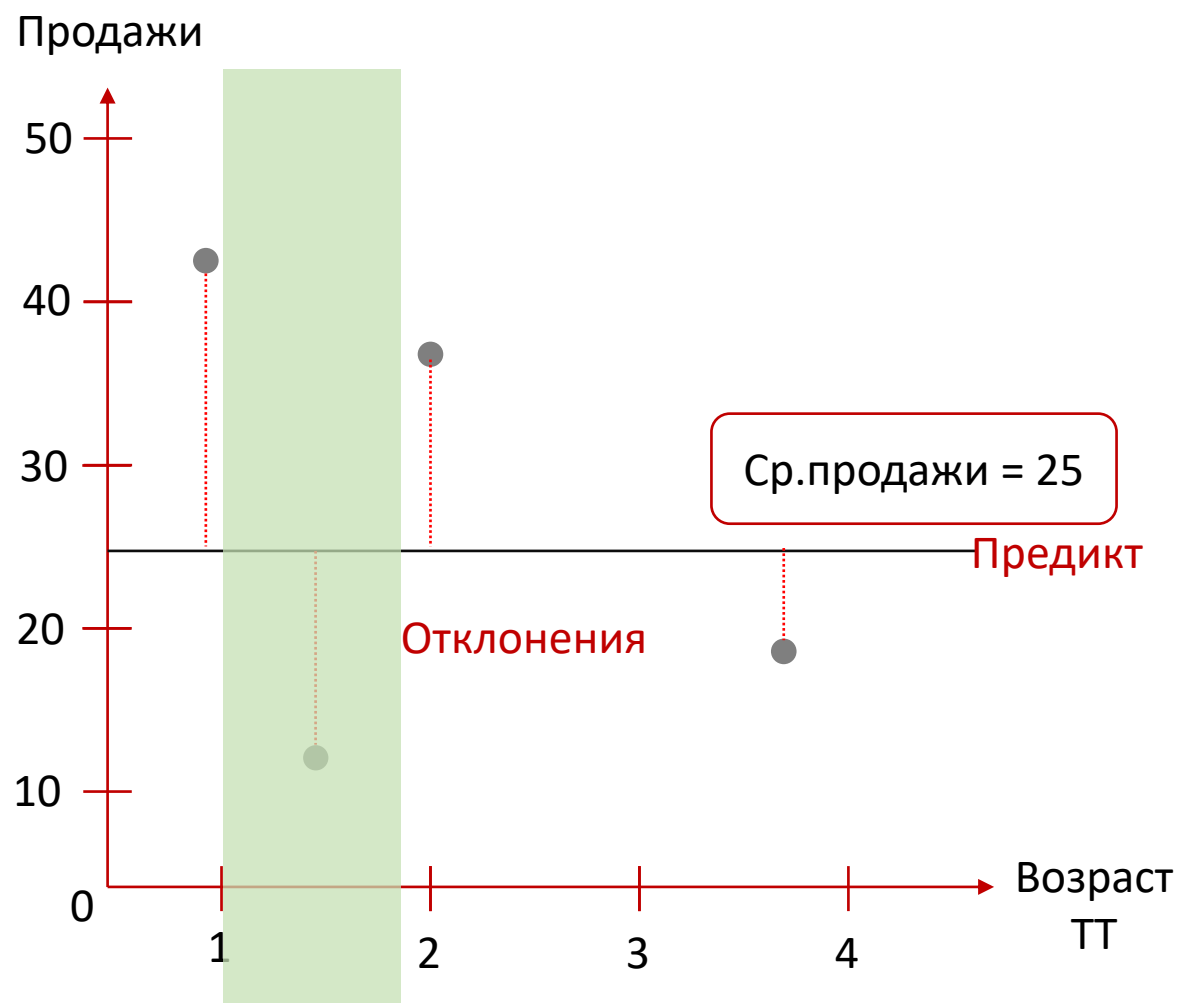
ПЛАН

1. Вспомним что было на Регрессии
2. Узнаем о Классификации

ЧТО БЫЛО ДО



TREE, GAIN, PRUNE



Gamma для Pruning? Случайное число

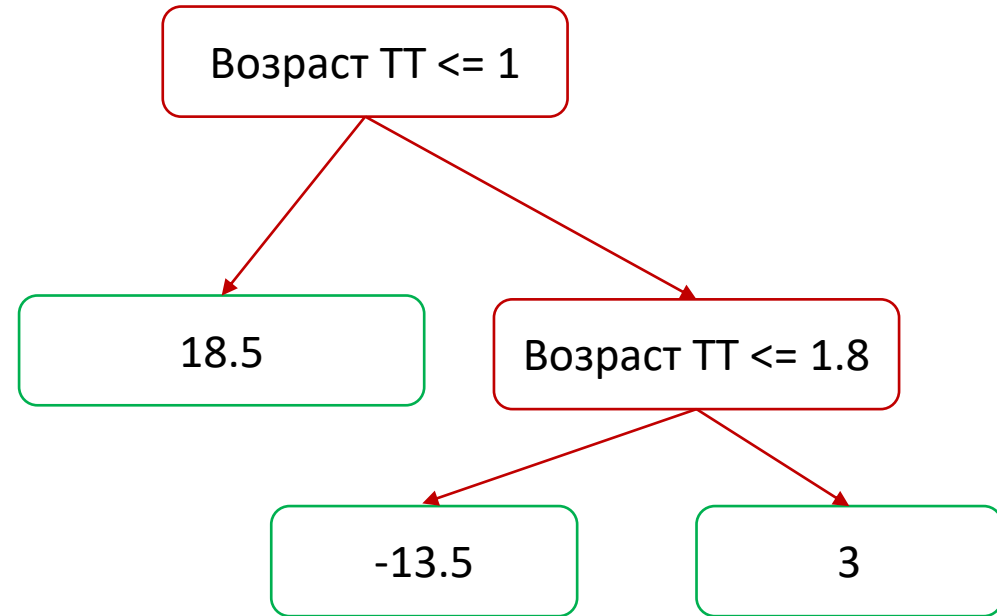
1. $GAIN - Gamma = 330.75 - 200 = 130.75$
2. $GAIN - Gamma = 181.5 - 200 = -18.5$

УЧИМСЯ НА DELTA

Возраст ТТ	Продажи	Delta
1,00	44,00	-18.5
1,50	12,00	13.5
2,00	38,00	-12.5
3,80	19,00	6.5



Минимизируем отклонение
факт-предик



$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$

$F(x) =$
25.5
 $+ 0.1 *$

Возраст ТТ <= 1

18.5

Возраст ТТ <= 1.8

-13.5

3

 $= 27.35$

Возраст ТТ	Продажи	Delta	Predict
1,00	44,00	18.5	27.35
1,50	12,00	-13.5	26.85
2,00	38,00	12.5	26.1
3,80	19,00	-6.5	26.1

The background of the slide is a solid red color. In the upper right, there is a large, semi-transparent orange sun with horizontal rays. Several palm trees are scattered across the middle ground, also semi-transparent. A faint grid of thin white lines covers the lower half of the slide.

ПРО КЛАССИФИКАЦИЮ

ПЕРВЫЙ ШАГ. НА ЧТО УЧИТЬСЯ?

$$Set(x_i, x_j)_{i=1}^n$$

Возраст ТТ	Была покупка
1,00	1
1,50	0
2,00	0
3,80	0

Функция предикта

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$

Функция потерь

$$\frac{1}{2}(\text{Observed} - \text{Predicted})^2 = \frac{2}{2}(\text{Observed} - \text{Predicted})x - 1 = -(\text{Observed} - \text{Predicted})$$

ПЕРВЫЙ ШАГ. НА ЧТО УЧИТЬСЯ?

$$Set(x_i, x_j)_{i=1}^n$$

Возраст ТТ	Была покупка
1,00	1
1,50	0
2,00	0
3,80	0

1 – положительный

3 - отрицательных

$$= 1/3 = 0.33$$

ПЕРВЫЙ ШАГ. НА ЧТО УЧИТЬСЯ?

$$Set(x_i, x_j)_{i=1}^n$$

Возраст ТТ	Была покупка
1,00	1
1,50	0
2,00	0
3,80	0

1 – положительный

3 - отрицательных

$$= 1/3 = 0.33 = \log(0.33) = -0.47$$

$$P = \text{LogReg}(\log(1/3)) = 0.38$$

$$\text{Log}(\text{Все } 1 / \text{Все } 0)$$

ПЕРВЫЙ ШАГ. НА ЧТО УЧИТЬСЯ?

$$Set(x_i, x_j)_{i=1}^n$$

Возраст ТТ	Была покупка
1,00	1
1,50	0
2,00	0
3,80	0

1 – положительный

3 – отрицательных

$$= 1/3 = 0.33 = \log(0.33) = -0.47$$

$$P = \text{LogReg}(\log(1/3)) = 0.38$$

$\log(\text{Всё } 1 / \text{Всё } 0)$



Classification
Threshold

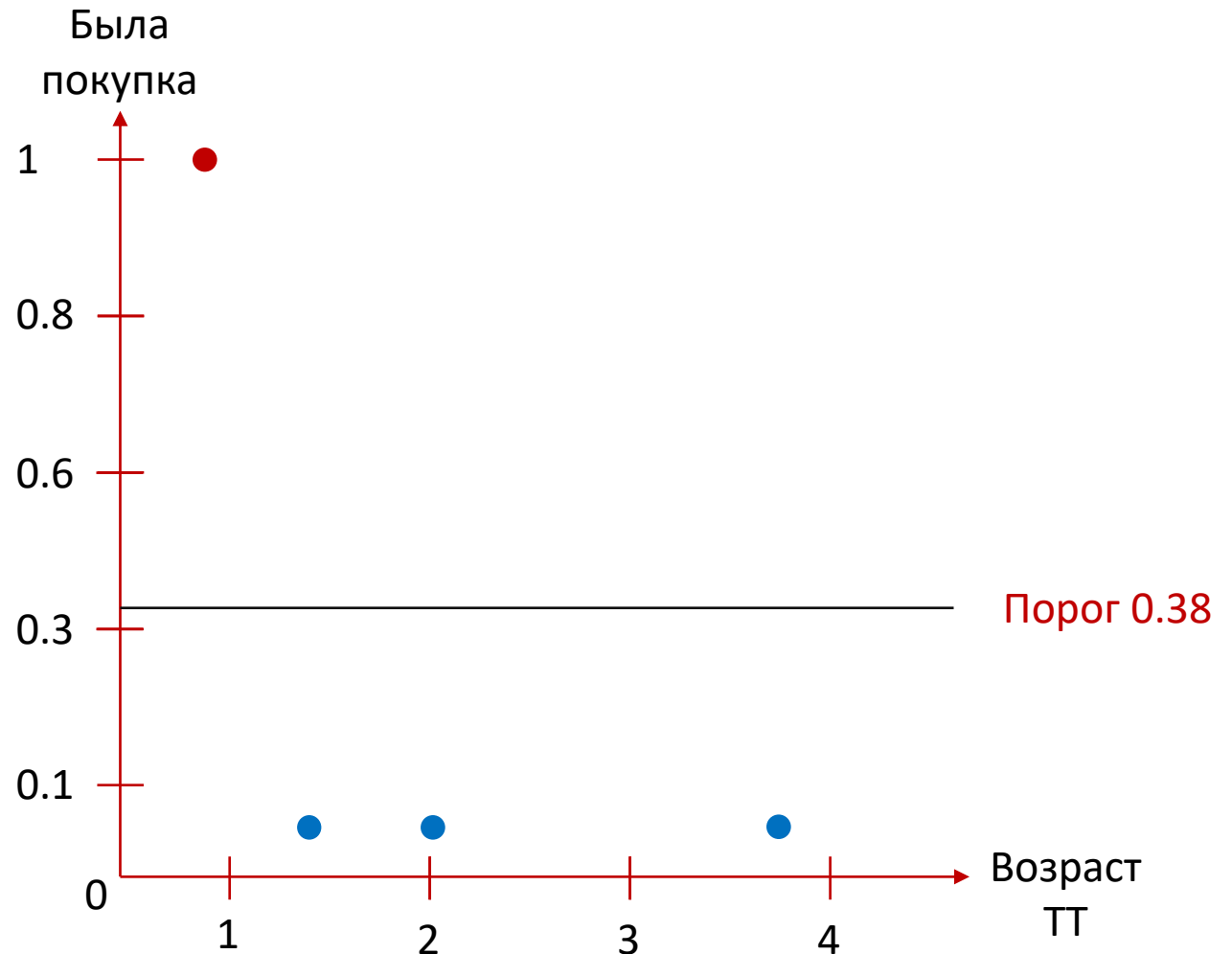
ПЕРВЫЙ ШАГ. НА ЧТО УЧИТЬСЯ?

$$Set(x_i, x_j)_{i=1}^n$$

Возраст ТТ	Была покупка
1,00	1
1,50	0
2,00	0
3,80	0

$$= 1/3 = 0.33 = \log(0.33) = -0.47$$

$$P = \text{LogReg}(\log(1/3)) = 0.38$$



ПЕРВЫЙ ШАГ. НА ЧТО УЧИТЬСЯ?

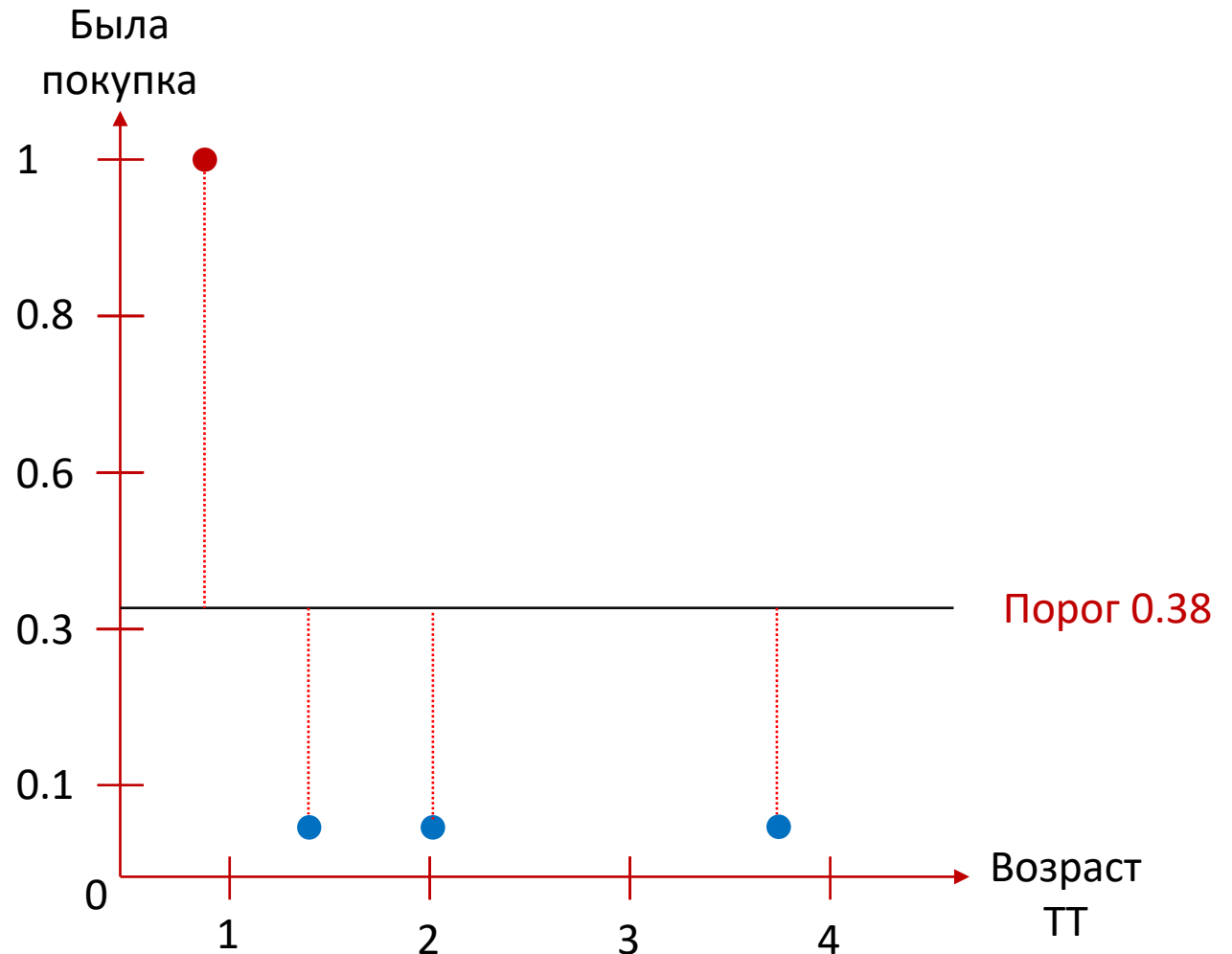
$$Set(x_i, x_j)_{i=1}^n$$

Возраст ТТ	Была покупка
1,00	1
1,50	0
2,00	0
3,80	0

$$P = 0.38$$

$$P \text{ возраст ТТ } 1 = 1 - 0.38 = 0.62$$

$$P \text{ возраст ТТ } 1.5 = 0 - 0.38 = -0.38$$



ПЕРВЫЙ ШАГ. УЧИМСЯ НА РАЗНИЦЕ

Возраст ТТ	Была покупка	Delta
1,00	1	0.62
1,50	0	-0.38
2,00	0	-0.38
3,80	0	-0.38

$P = 0.38$

$P \text{ возраст ТТ } 1 = 1 - 0.38 = 0.62$

$P \text{ возраст ТТ } 1.5 = 0 - 0.38 = -0.38$

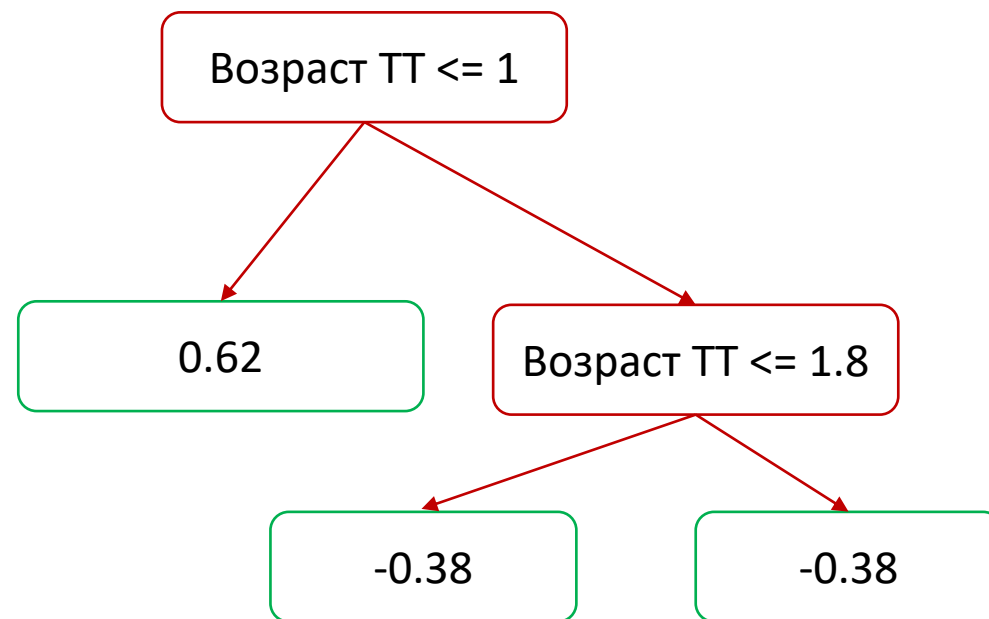
ПЕРВЫЙ ШАГ. УЧИМСЯ НА РАЗНИЦЕ

Возраст ТТ	Была покупка	Delta
1,00	1	0.62
1,50	0	-0.38
2,00	0	-0.38
3,80	0	-0.38

$P = 0.38$

P **возраст ТТ 1** = $1 - 0.38 = 0.62$

P **возраст ТТ 1.5** = $0 - 0.38 = -0.38$



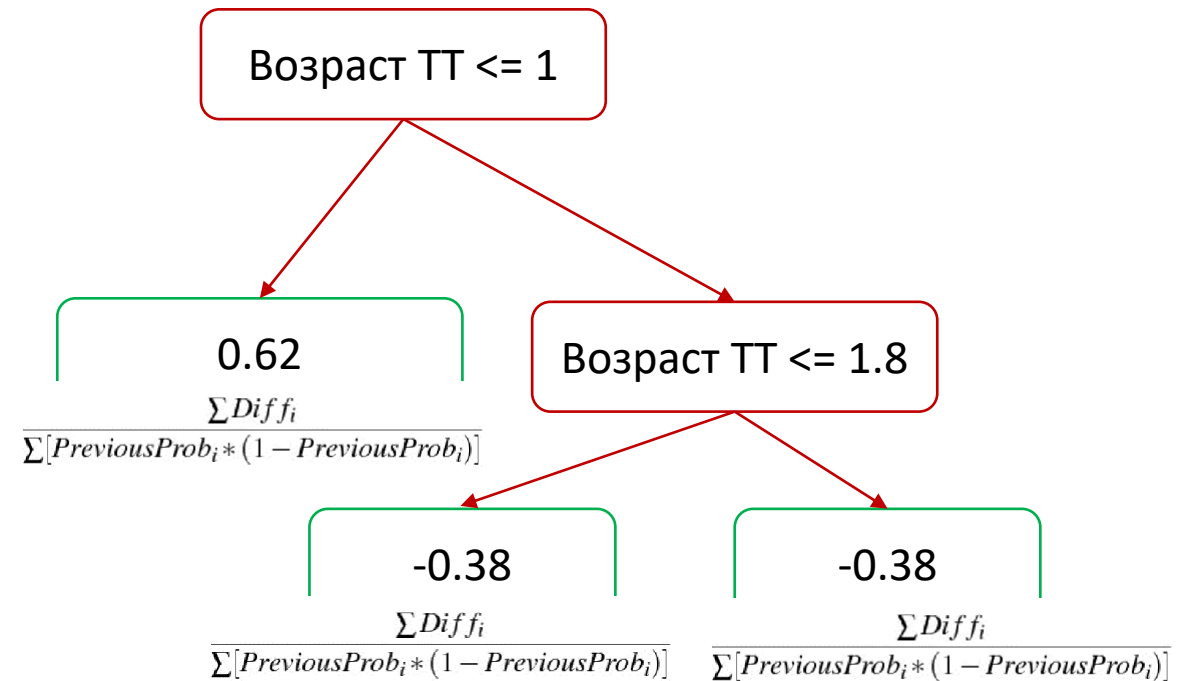
ИЗМЕНЯЕМ РЕЗУЛЬТАТ

Возраст ТТ	Была покупка	Delta
1,00	1	0.62
1,50	0	-0.38
2,00	0	-0.38
3,80	0	-0.38

$P = 0.38$

P **возраст ТТ 1** = $1 - 0.38 = 0.62$

P **возраст ТТ 1.5** = $0 - 0.38 = -0.38$



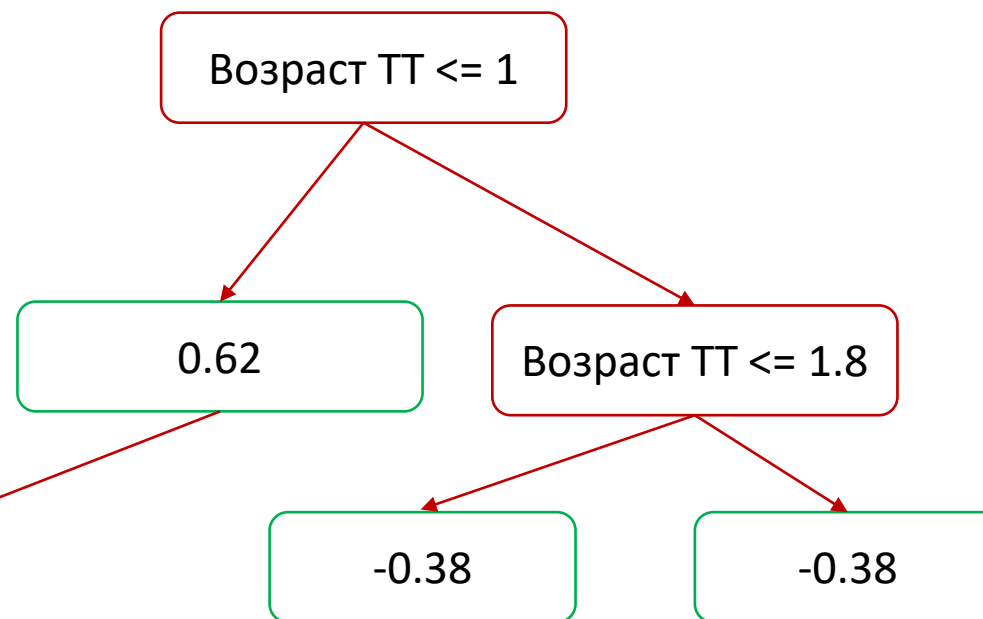
ИЗМЕНЯЕМ РЕЗУЛЬТАТ

Возраст ТТ	Была покупка	Delta
1,00	1	0.62
1,50	0	-0.38
2,00	0	-0.38
3,80	0	-0.38

P = 0.38

$$\frac{\sum Diff_i}{\sum [PreviousProb_i * (1 - PreviousProb_i)]}$$

$$= 0.62 / (0.38 * (1 - 0.38)) = 2.63$$



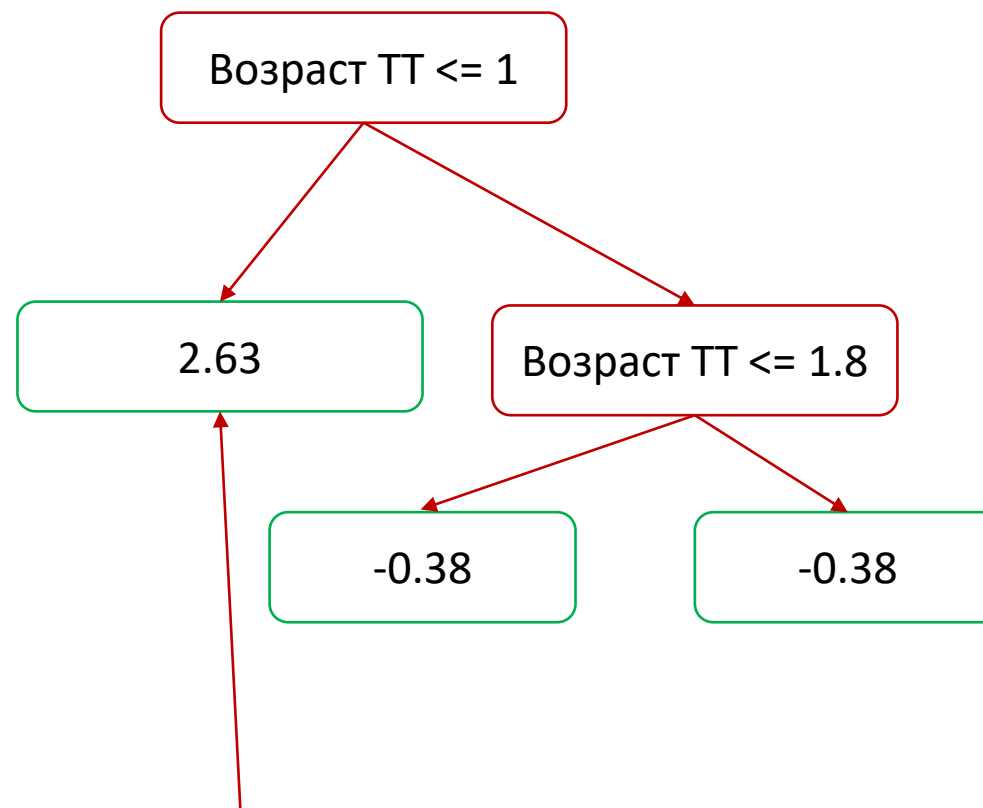
ИЗМЕНЯЕМ РЕЗУЛЬТАТ

Возраст ТТ	Была покупка	Delta
1,00	1	0.62
1,50	0	-0.38
2,00	0	-0.38
3,80	0	-0.38

P = 0.38

$$\frac{\sum Diff_i}{\sum [PreviousProb_i * (1 - PreviousProb_i)]}$$

$$= 0.62 / (0.38 * (1 - 0.38)) = 2.63$$



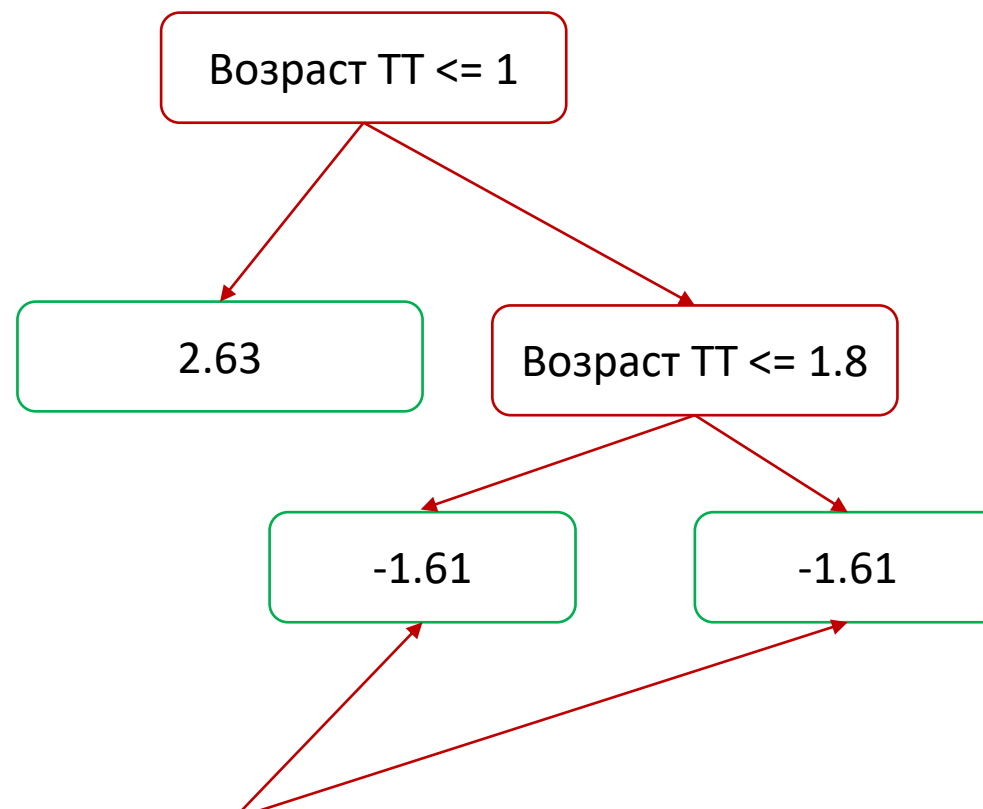
ИЗМЕНЯЕМ РЕЗУЛЬТАТ

Возраст ТТ	Была покупка	Delta
1,00	1	0.62
1,50	0	-0.38
2,00	0	-0.38
3,80	0	-0.38

P = 0.38

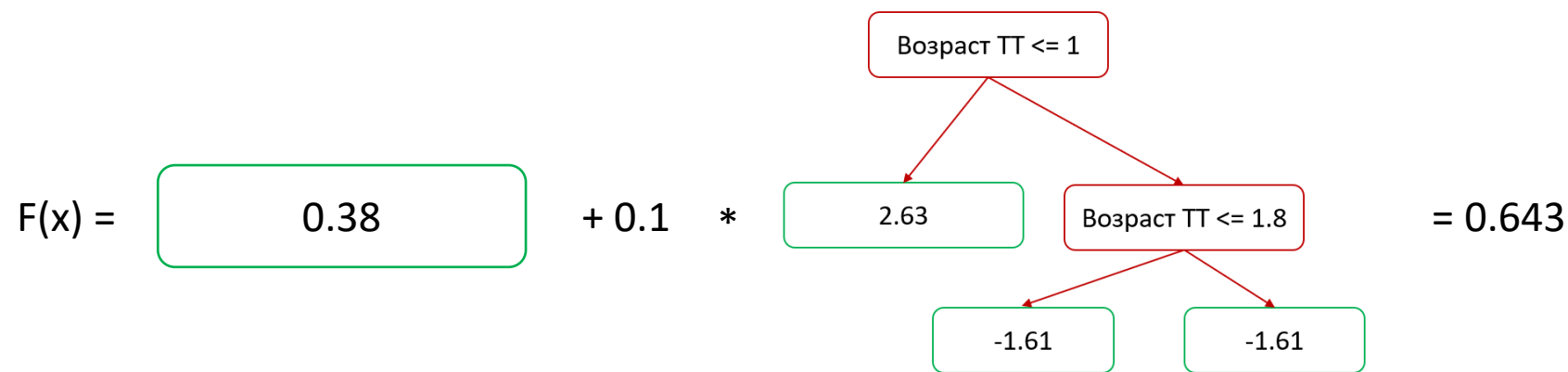
$$\frac{\sum Dif f_i}{\sum [PreviousProb_i * (1 - PreviousProb_i)]}$$

$$= -0.38 / (0.38 * (1 - 0.38)) = 2.63$$



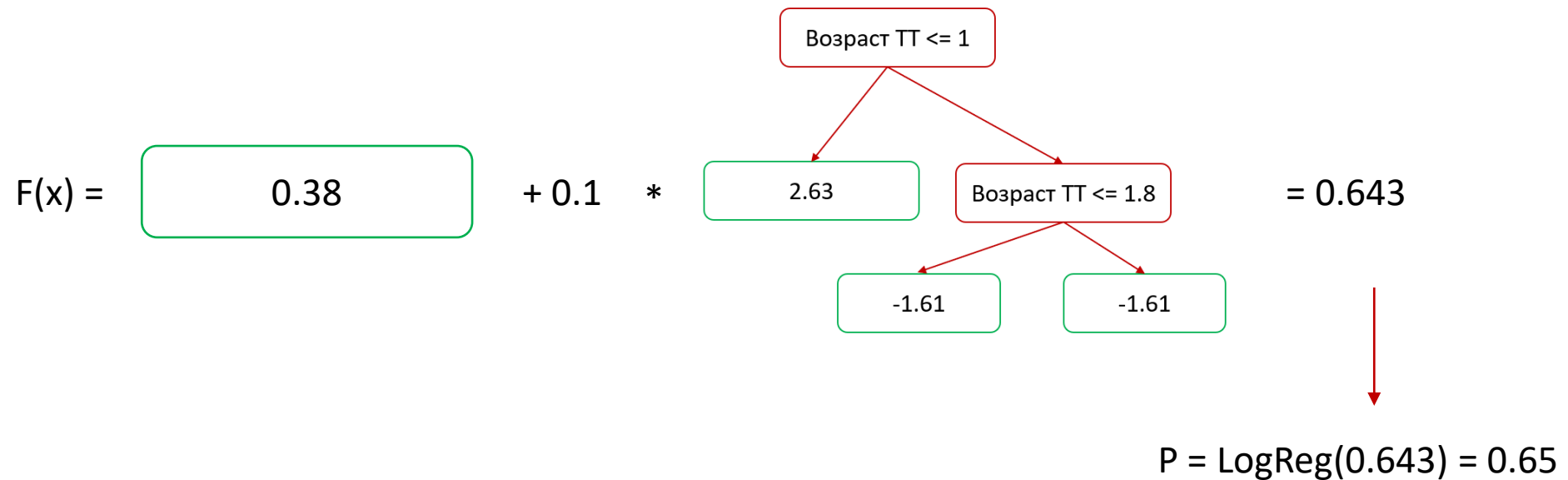
ПЕРВЫЙ ПРЕДИКТ

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$



ПЕРВЫЙ ПРЕДИКТ

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$



ПЕРВЫЙ ПРЕДИКТ

Возраст ТТ	Была покупка	Delta	Prob
1,00	1	0.62	P = LogReg(0.643) = 0.65 0.65
1,50	0	-0.38	
2,00	0	-0.38	
3,80	0	-0.38	

ПЕРВЫЙ ПРЕДИКТ

Возраст ТТ	Была покупка	Delta		Prob
1,00	1	0.62	$P = \text{LogReg}(0.643) = 0.65$	0.65
1,50	0	-0.38	$P = \text{LogReg}(0.219) = 0.55$	0.55
2,00	0	-0.38	$P = \text{LogReg}(0.219) = 0.55$	0.55
3,80	0	-0.38	$P = \text{LogReg}(0.219) = 0.55$	0.55

НОВАЯ ДЕЛЬТА

Возраст ТТ	Была покупка	Prob	Delta
1,00	1	0.65	0.35
1,50	0	0.55	-0.45
2,00	0	0.55	-0.45
3,80	0	0.55	-0.45

ПАРУ СЛОВ ПРО ОТЛИЧИЯ



ОСНОВНЫЕ ОТЛИЧИЯ

1. Использование Likelihood в функции потерь

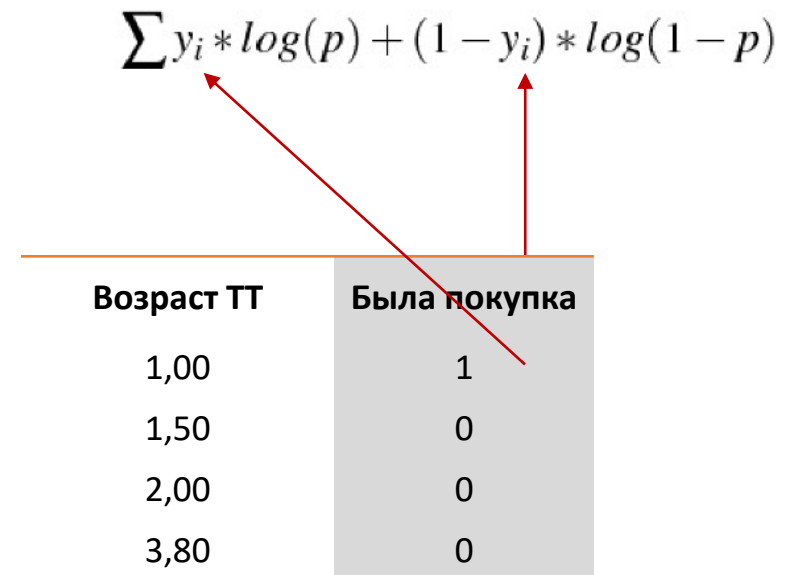
$$\sum y_i * \log(p) + (1 - y_i) * \log(1 - p)$$

$$P = \text{LogReg}(\text{Log}(1/3)) = 0.38$$
A diagram consisting of two red arrows. One arrow originates from the 'P' in the equation below and points upwards and to the left towards the 'p' in the first term of the equation above. The second arrow originates from the same point and points upwards and to the right towards the 'p' in the second term of the equation above.

ОСНОВНЫЕ ОТЛИЧИЯ

1. Использование Likelihood в функции потерь

$$\sum y_i * \log(p) + (1 - y_i) * \log(1 - p)$$



Возраст ТТ	Была покупка
1,00	1
1,50	0
2,00	0
3,80	0

ОСНОВНЫЕ ОТЛИЧИЯ

1. Использование Likelihood в функции потерь
2. Другое вычисление R на листьях
Через LogReg

$$r_{i,m} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$



$$-Observed + \frac{e^{\log(odds)}}{1 + e^{\log(odds)}}$$

ОСНОВНЫЕ ОТЛИЧИЯ

1. Использование Likelihood в функции потерь
2. Другое вычисление R на листьях
Через LogReg
3. Другое вычисление Гаммы

$$\gamma = \frac{-\frac{\partial}{\partial F()}(y_i, F_{m-1}(x_i))}{\frac{\partial^2}{\partial F()^2}(y_i, F_{m-1}(x_i))}$$

Факт - Предикт

↑

↓

Предикт * (1 - Предикт)