

# Оценка качества алгоритмов машинного обучения. Кросс- валидация.

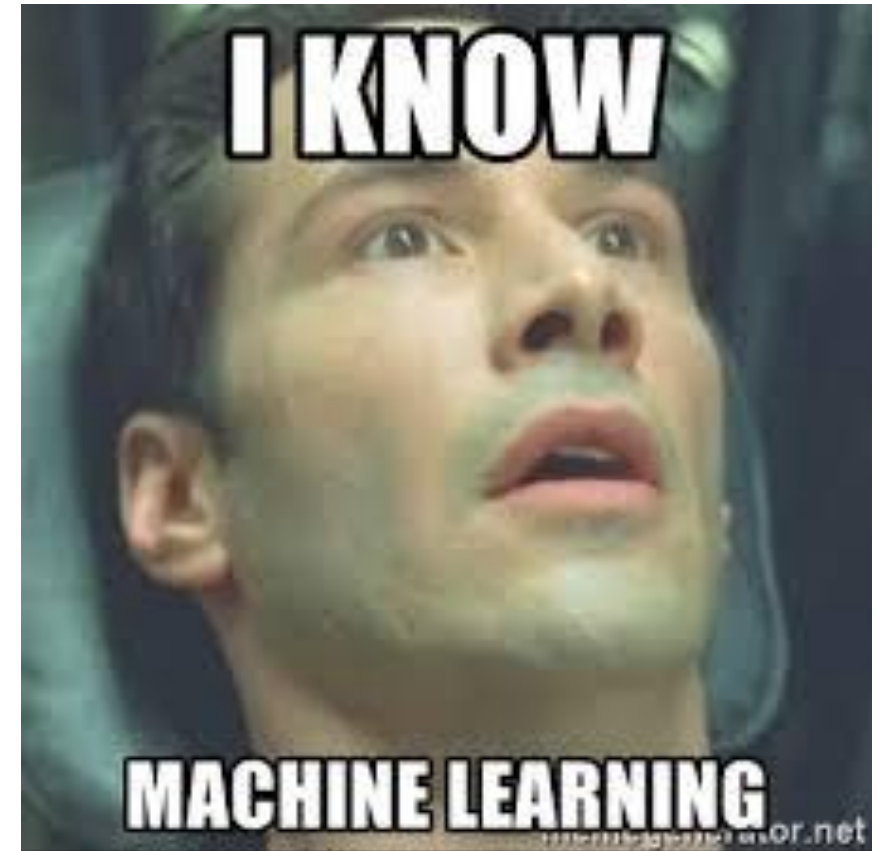
Екатерина Кондратьева  
Полина Полунина

```
from sklearn.linear_model import LinearRegression
```

```
lr = LinearRegression()
```

```
lr.fit(X_train,y_train)
```

```
lr.predict(X_test)
```



**MUCH TO LEARN YOU STILL HAVE**

**MY YOUNG PADAWAN**

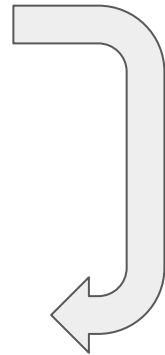
quickmeme.com

# Пайплайн валидации модели:

1. Выбор метрики валидации
2. Валидация и выбор модели:  
валидация, чаще кросс валидация(выбор модели)
3. Тестирование лучшей модели

# Пайплайн валидации модели:

1. Выбор метрики валидации
2. Выбор модели:  
валидация (чаще кросс - валидация)
3. Тестирование лучшей модели



# Пайплайн валидации модели:

1. Выбор метрики валидации
2. Выбор модели:  
валидация (чаще кросс - валидация)
3. Тестирование лучшей модели

Много чисел, сравнение  
распределений

Одно значение, финальная  
характеристика модели

# Метрики

# Оценка качества алгоритмов машинного обучения с учителем

## Классификация

Accuracy

TP Rate

TN Rate

FP Rate

FN Rate

*\*Confusion matrix*

Precision

Recall/Sensitivity

Specificity

ROC\_AUC

PR\_AUC

F1 Score

F-Beta Score

Log Loss...

## Регрессия

MAE

MAPE

MSE

RMSE

$R^2$

$R^2$  adjusted

## Кластерный анализ, мультикласс



Без учителя?

# Метрики для задачи классификации: Confusion Matrix

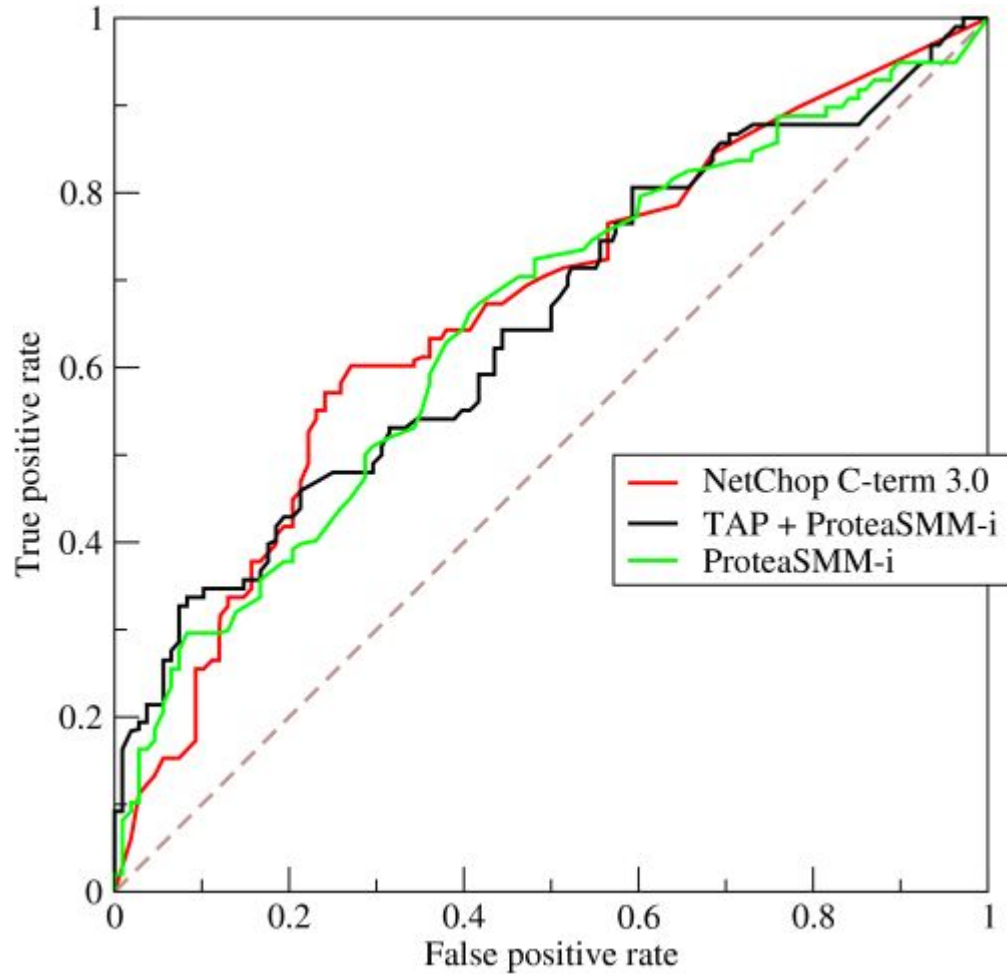
n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) <b>Type II Error</b>	<b>Sensitivity</b> $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) <b>Type I Error</b>	True Negative (TN)	<b>Specificity</b> $\frac{TN}{(TN + FP)}$
		<b>Precision</b> $\frac{TP}{(TP + FP)}$	<b>Negative Predictive Value</b> $\frac{TN}{(TN + FN)}$	<b>Accuracy</b> $\frac{TP + TN}{(TP + TN + FP + FN)}$

# Метрики для задачи классификации:

- F1 Score =  $2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$
- $$F_{\beta} = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}}$$
- ROC\_AUC - площадь под ROC кривой
- PR\_AUC – площадь под Precision-Recall кривод

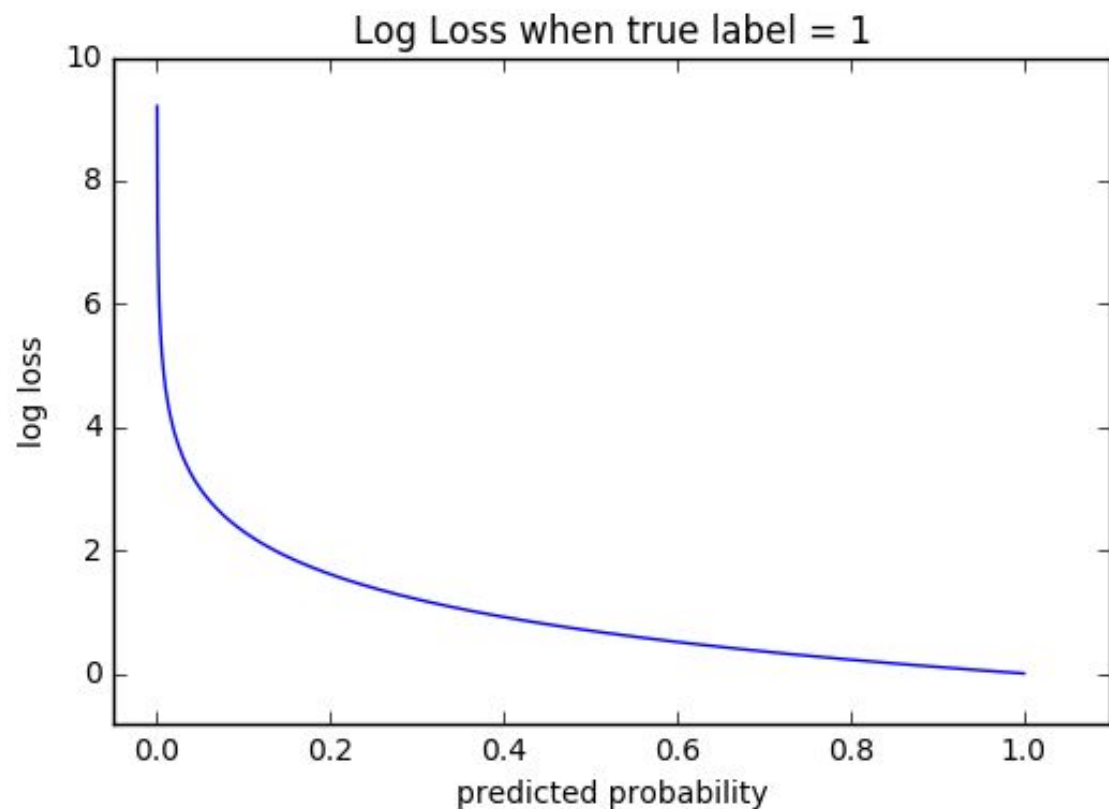
# Метрики для задачи классификации: ROC кривая



Двигаем ползунок порога, при котором объект причисляется к классу 0 или 1.

Какой дефолтный порог?

# Метрики для задачи классификации: Log Loss



```
def logloss(true_label, predicted, eps=1e-15):  
    p = np.clip(predicted, eps, 1 - eps)  
    if true_label == 1:  
        return -log(p)  
    else:  
        return -log(1 - p)
```

# Метрики для задачи регрессии:

- $$\text{MAE} = \frac{\sum_{i=1}^n |y_i - x_i|}{n} = \frac{\sum_{i=1}^n |e_i|}{n}$$
- $$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$
- $$\text{RMSE} = \sqrt{\text{MSE}}$$
- $$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

MAE или RMSE более устойчива к выбросам?

# Метрики для задачи регрессии: RMSE

CASE 1: Evenly distributed errors

ID	Error	Error	Error^2
1	2	2	4
2	2	2	4
3	2	2	4
4	2	2	4
5	2	2	4
6	2	2	4
7	2	2	4
8	2	2	4
9	2	2	4
10	2	2	4

MAE	RMSE
2.000	2.000

CASE 2: Small variance in errors

ID	Error	Error	Error^2
1	1	1	1
2	1	1	1
3	1	1	1
4	1	1	1
5	1	1	1
6	3	3	9
7	3	3	9
8	3	3	9
9	3	3	9
10	3	3	9

MAE	RMSE
2.000	2.236

CASE 3: Large error outlier

ID	Error	Error	Error^2
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	0	0	0
9	0	0	0
10	20	20	400

MAE	RMSE
2.000	6.325



# Метрики для задачи регрессии: R^2

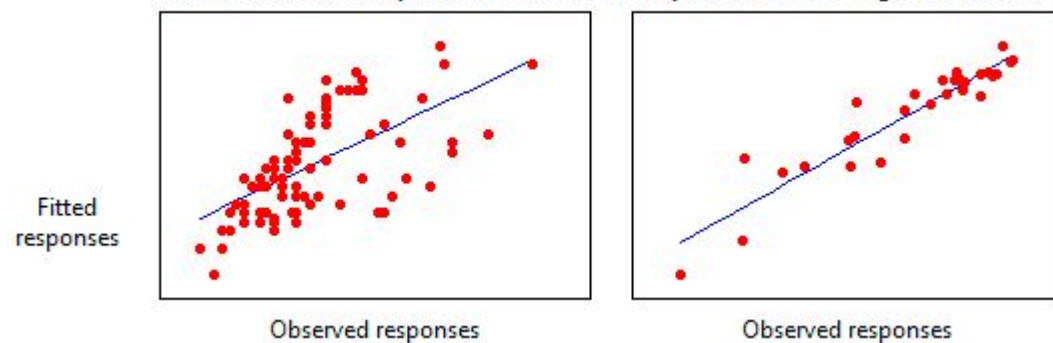
$$R^2 = 1 - \frac{D[y|x]}{D[y]} = 1 - \frac{\sigma^2}{\sigma_y^2},$$

дисперсия ошибки модели  
дисперсия случайной величины y

R^2 (Coefficient of determination) - это доля дисперсии зависимой переменной, объясняемая рассматриваемой моделью зависимости, то есть объясняющими переменными.

При оценке регрессионных моделей это интерпретируется как соответствие модели данным.

Plots of Observed Responses Versus Fitted Responses for Two Regression Models



R2 = 38%

R2 = 87%

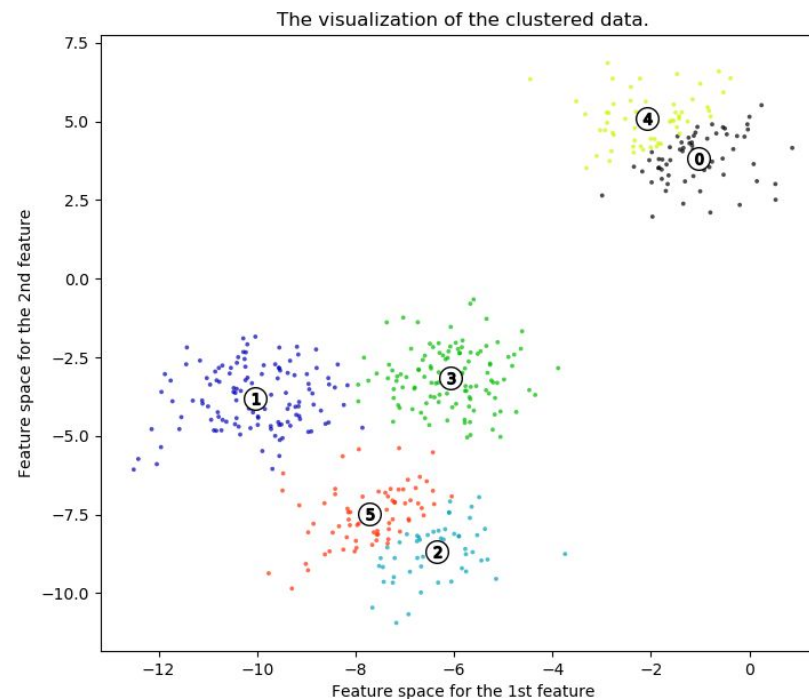
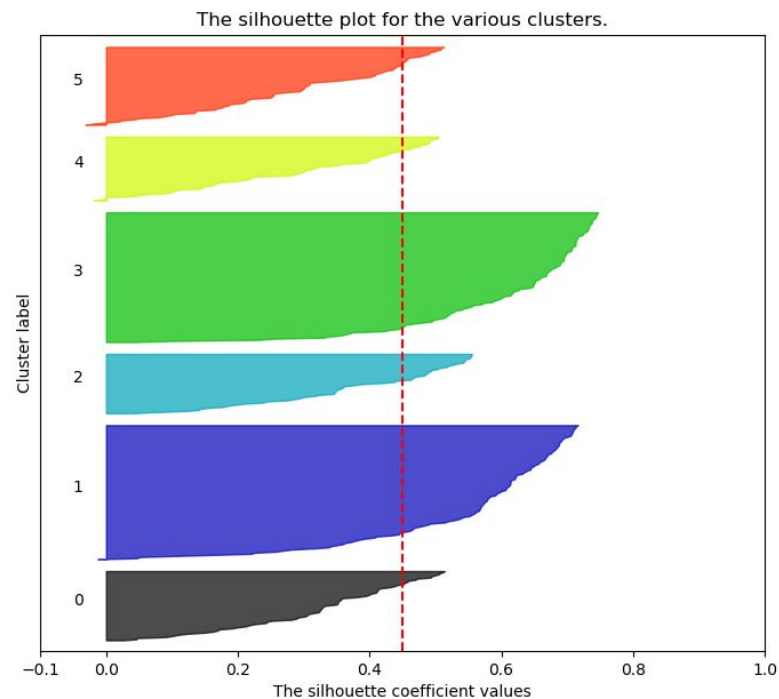
Может ли R2 быть отрицательным?

можно ли оценить регрессионную модель в  
точности (accuracy)?

# Кластерный анализ и мультиномиальная регрессия

# Кластерный анализ

Silhouette analysis for KMeans clustering on sample data with  $n\_clusters = 6$



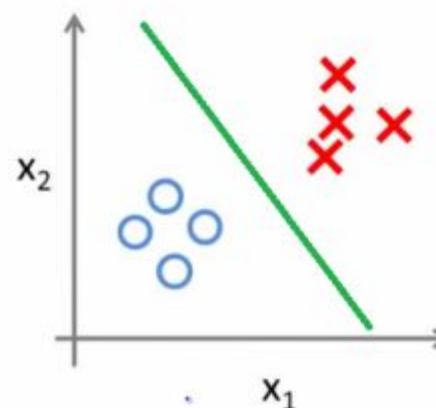
# Мультиклассовая классификация

## One-vs-rest or One-vs-One:

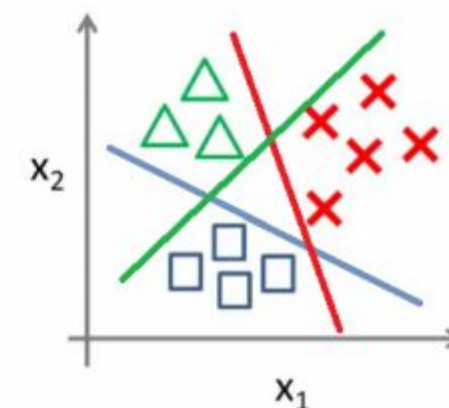
Тогда для нового объекта выборки решение будет принято в соответствии с наибольшим скором одного из классификаторов:

$$\hat{y} = \operatorname{argmax}_{k \in \{1 \dots K\}} f_k(x)$$

Binary classification:

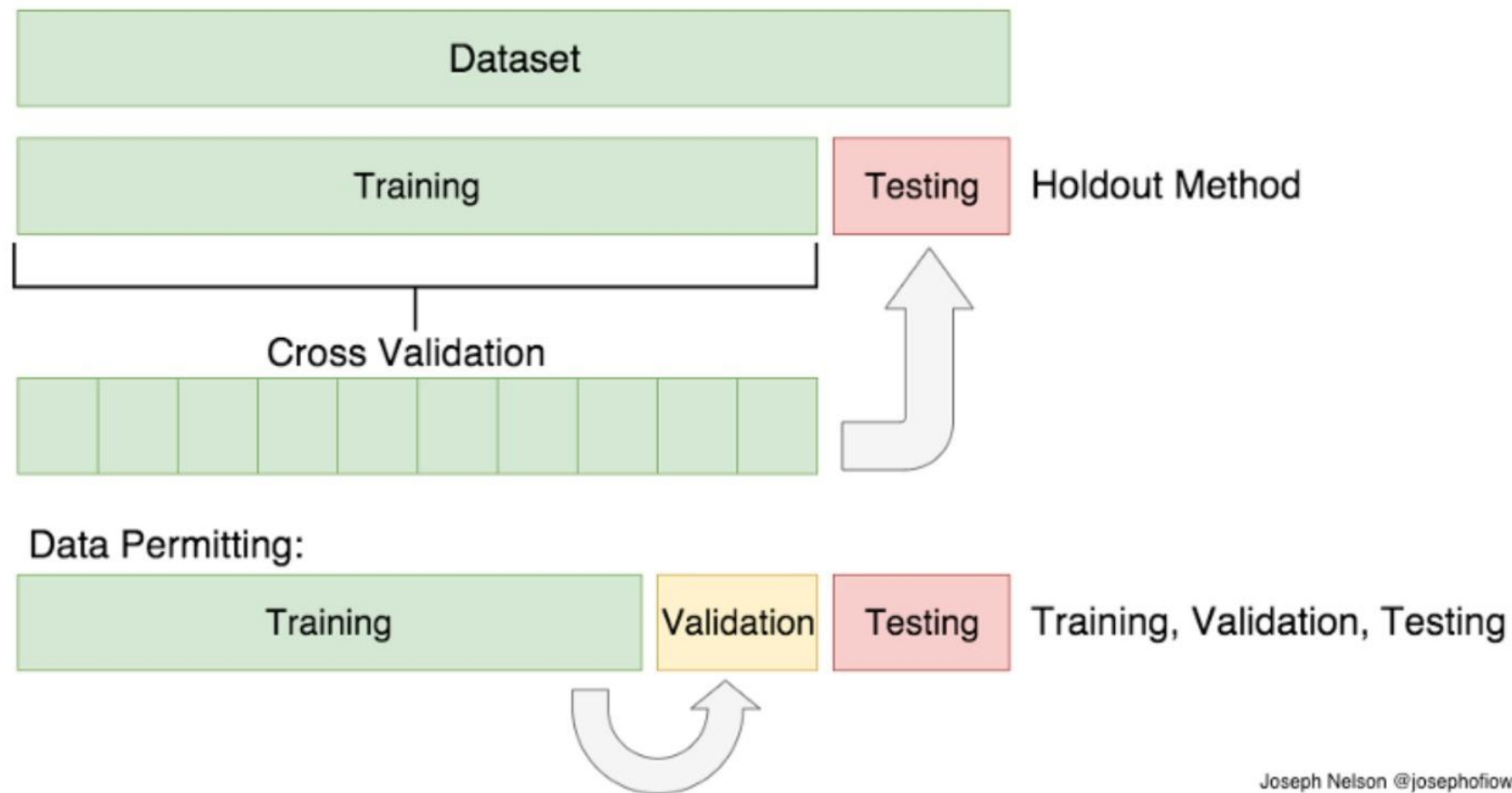


Multi-class classification:



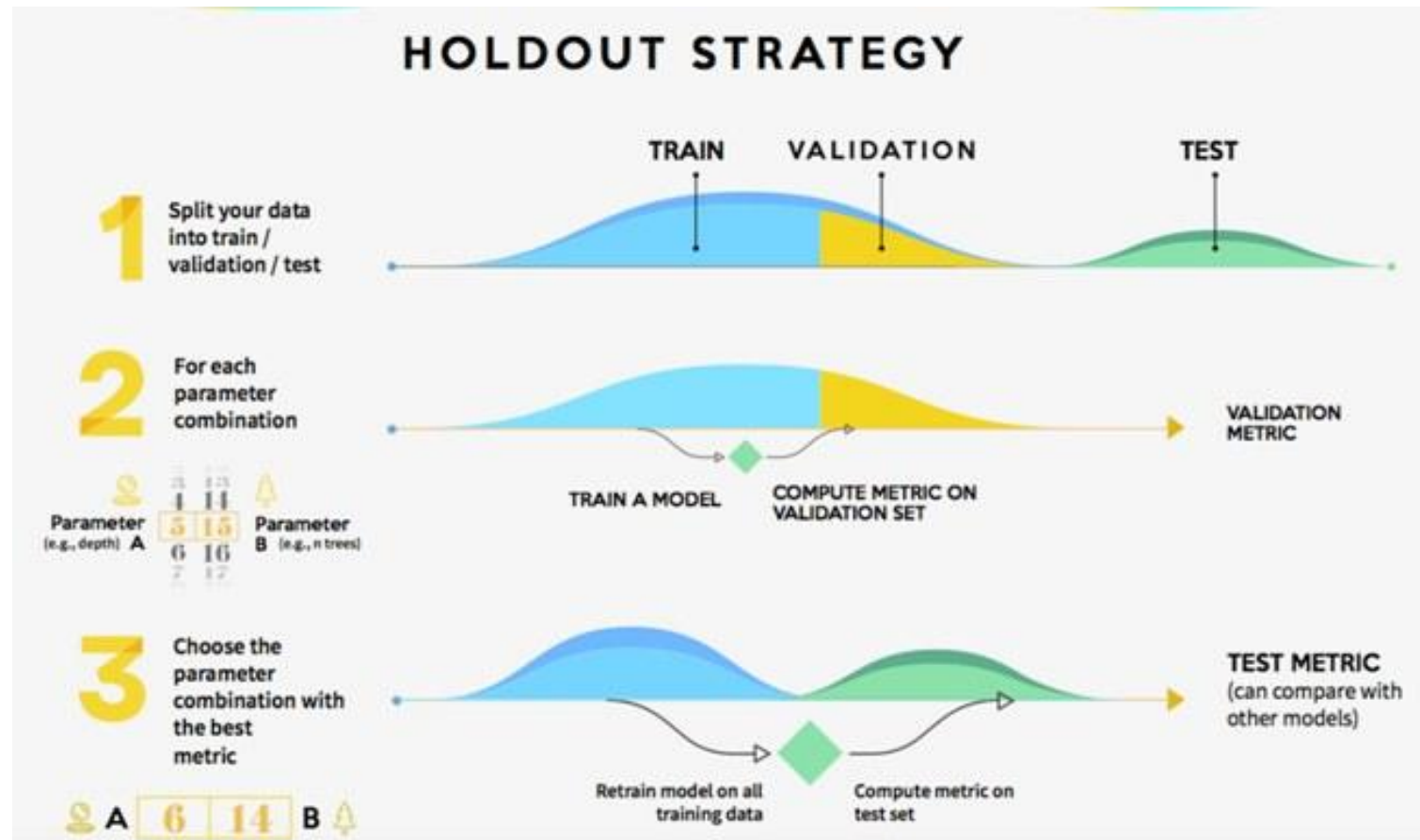
# Разделение данных

# Подход к разделению данных на подвыборки



# Подход к разделению данных на подвыборки

Основная идея: разделить выборку на несколько независимых частей, чтобы оценить обобщающую способность модели





# Методы валидации:

- **Train/Test Split:** С одной стороны,  $k$  может быть равно 1, т.е. всего одно разделение на train/test
- **LOOCV:** С другой стороны,  $k$  может быть равно кол-ву наблюдений в датасете, т.е. предсказание делается каждый раз на одном наблюдении. Такой подход называется leave-one-out cross-validation
- **Repeated:** Разбиваем на  $k$  фолдов несколько раз (KFold)
  - \***Stratified:** В задачах классификации с несбалансированными данными возникает потребность разбить данные на куски так, чтобы в каждом куске пропорция классов сохранялась.

# Методы валидации:

- **Train/Test Split:** С одной стороны, к может быть всего одно разделение на train/test
- **LOOCV:** С другой стороны, к может быть равно количеству наблюдений в датасете, т.е. предсказание делается каждое в отдельном наблюдении. Такой подход называется leave-one-out cross validation
- **Repeated:** Разбиваем на k фолдов несколько раз (KFold)
  - \* **Stratified:** В задачах классификации с несбалансированными данными возникает потребность разбить данные на куски так, чтобы в каждом куске пропорция классов сохранялась.

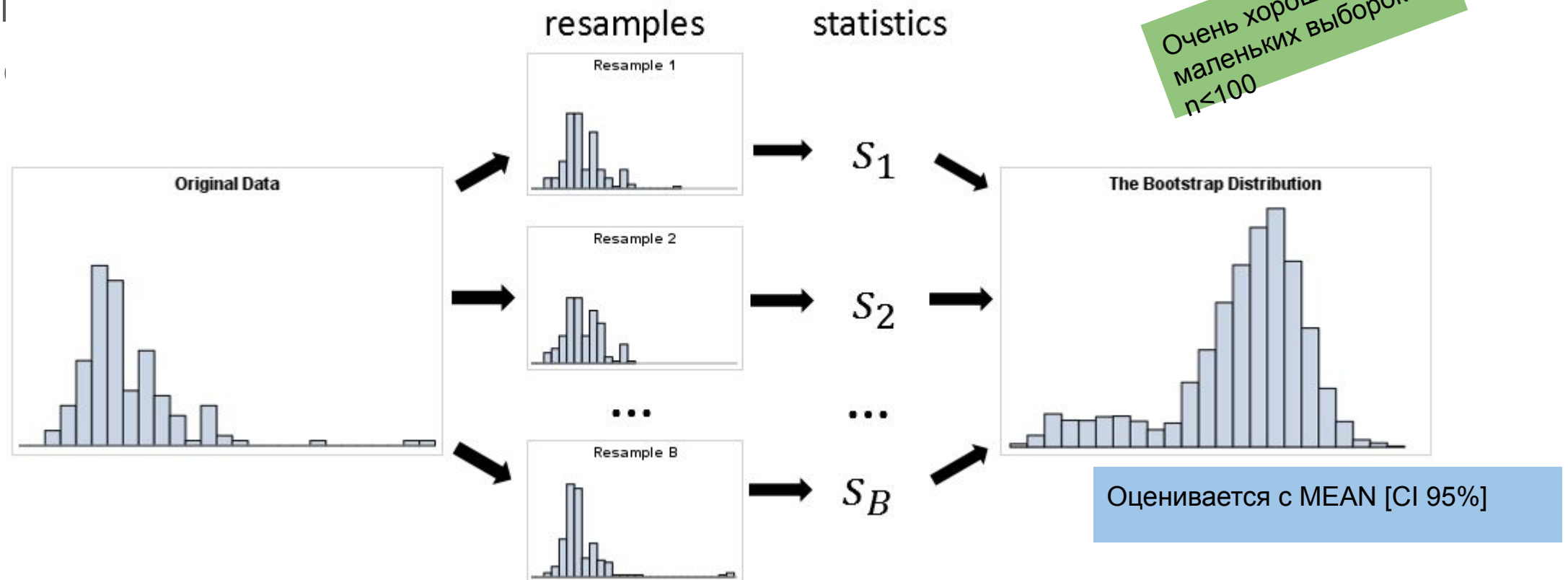
Смещенная оценка

Вероятность переобучиться, но часто требуется в специфической области

уже лучше

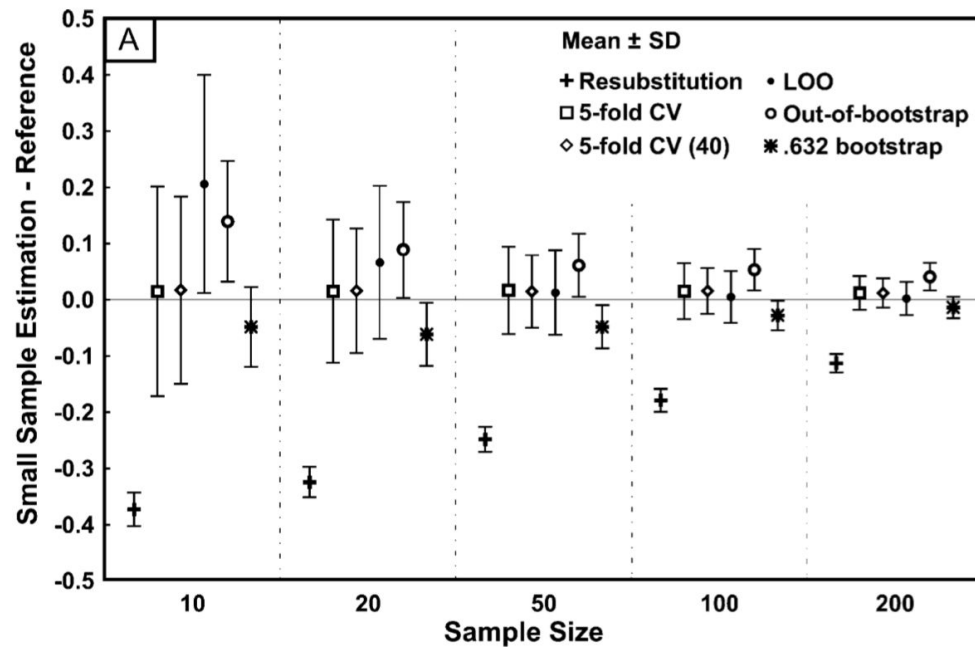
# Методы валидации: Bootstrap

- Train/Test Split. Оценить модель на тестовых данных.

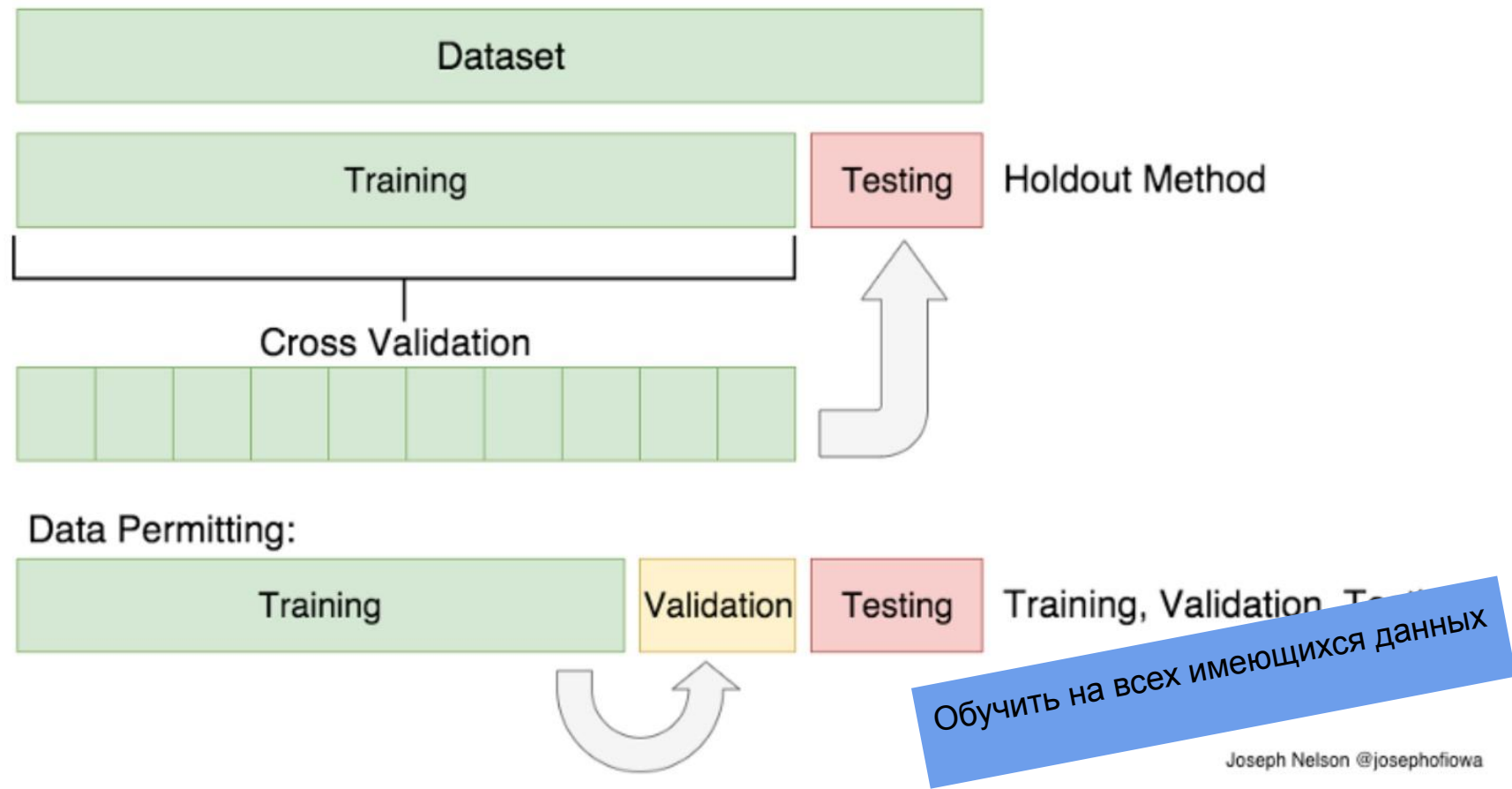


# Сколько нннааада фолдов?

*C. Beleites et al. / Chemometrics and Intelligent Laboratory Systems 79 (2005) 91–100*



# После тестирования модели:



Кейсы

# Пайплайн валидации модели:

Кейс: Классификация `sklearn.datasets.load_diabets` [442, 10]

1. Выбор метрики валидации
2. Разделение данных:  
валидация, чаще кросс валидация(выбор модели)
3. Тестирование лучшей модели

# Пайплайн валидации модели.

Кейс: Классификация `sklearn.datasets.load_breast_cancer` [569, 30]

1. Выбор метрики валидации  
ROC/AUC
2. Валидация и выбор модели:  
10 Fold CV5 with Stratification
3. Тестирование лучшей модели  
Предсказание рака груди с точностью 89%



# Пайплайн валидации модели.

Кейс: Классификация `sklearn.datasets.load_diabetes` [442, 10]

1. Выбор метрики валидации

MAE, RMSE,  $R^2$

2. Валидация и выбор модели:

10 Fold CV5

3. Тестирование лучшей модели

Предсказание степени тяжести диабета с точностью 89%  
и относительной ошибкой 5%

# Пайплайн валидации модели.

Кейс: Классификация `sklearn.datasets.load_diabets` [442, 10]

1. Выбор метрики валидации  
MAE, RMSE,  $R^2$
2. Валидация и выбор модели:  
10 Fold CV5
3. Тестирование лучшей модели

Предсказание степени тяжести диабета с точностью  
и относительной ошибкой 5%

Какие характеристики отвечают за это?

A polar bear is seen from the chest up, looking out of a window. Outside the window, a person wearing a bright yellow jacket and a hood is visible, looking towards the bear. The scene appears to be inside a building, possibly a zoo or a research facility.

**EXCUSE ME**

**DO YOU HAVE A MOMENT TO TALK ABOUT  
INTERPRETABLE MACHINE LEARNING**

memegenerator.net

# Пайплайн валидации модели.

Кейс: Классификация `sklearn.datasets.load_diabets` [442, 10]

1. Выбор метрики валидации

MAE, RMSE,  $R^2$

2. Валидация и выбор модели:

10 Fold CV5

3. Тестирование лучшей модели

Предсказание степени тяжести диабета с точностью  
и относительной ошибкой 5%

4. Интерпретация модели

Какие характеристики отвечают за это?

# Однозначная интерпретация результатов предсказания + к `feature_importance`

1. Корреляционный анализ
2. Дисперсионный анализ
3. Простейшие модели (Linear model, Decision Tree)

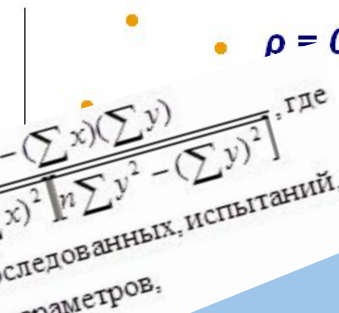
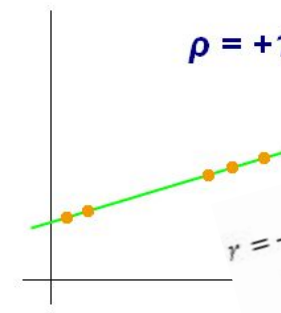
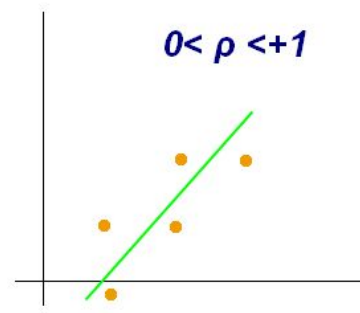
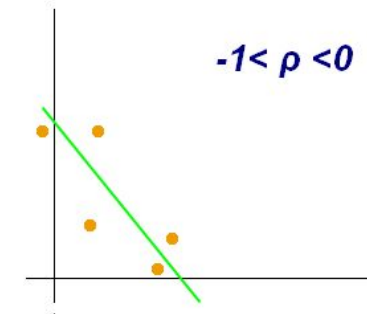
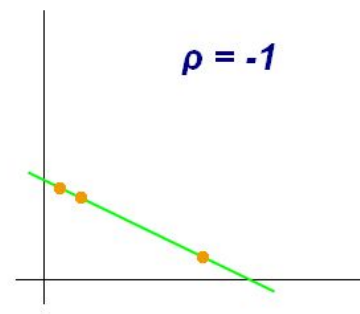
	<b>Categorical</b>	<b>Continuous</b>
<b>Categorical</b>	Lambda, Corrected Cramer's V	Point Biserial, Logistic Regression
<b>Continuous</b>	Point Biserial, Logistic Regression	Spearman, Kendall, Pearson

Один из вариантов

# Корреляционный анализ

## 1. Continuous and continuous variable

- Pearson corr



Проверяем линейную скоррелированность данных.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

где  
 $n$  - количество обследованных, испытаний,  
 $x$  - один из двух параметров,  
 $y$  - другой параметр,  
 $i$  - порядковы

знакомая R - статистика

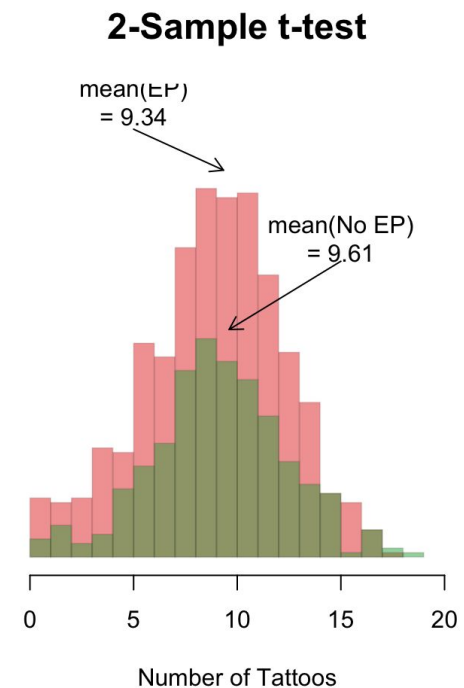
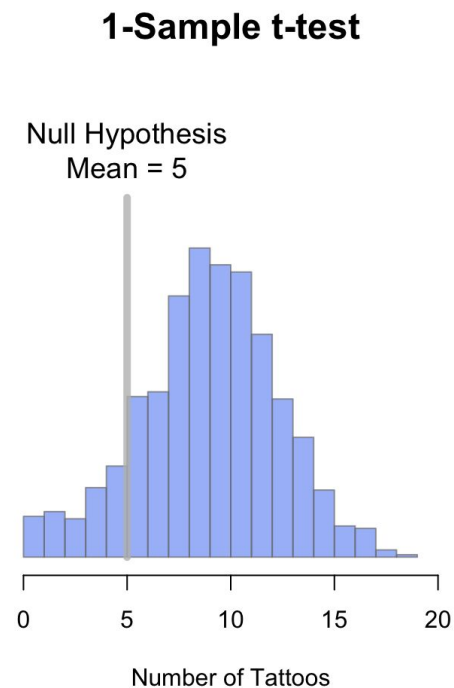
# Статистический тест

## 1. Continuous and continuous variable

- parametric t-test
- unpaired exact t-test

Проверяем гипотезу, что выборки из одного распределения.

2- Sample что значит?

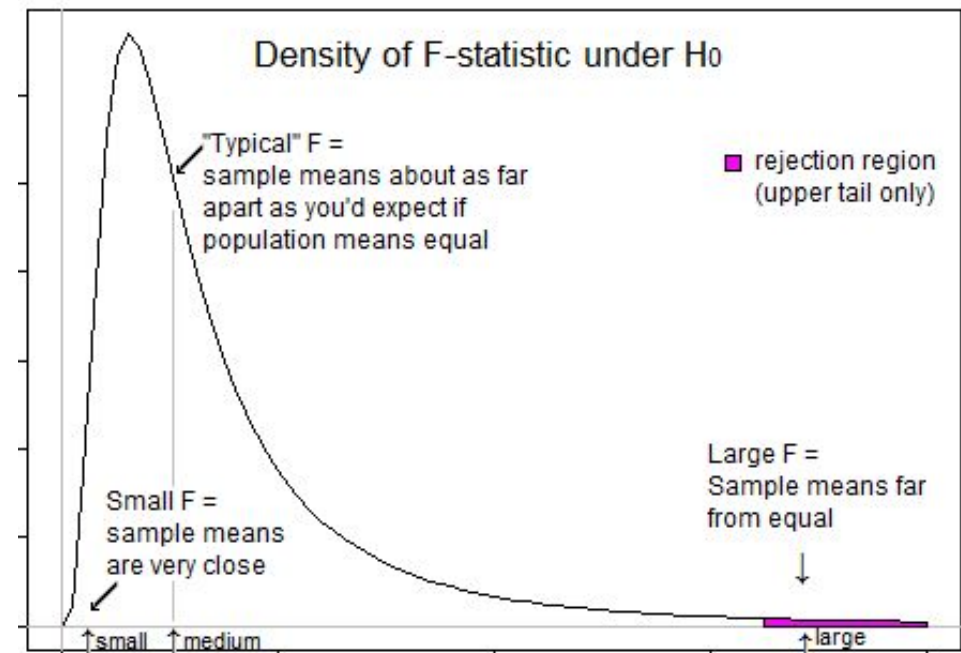


# Дисперсионный анализ

ANOVA (от **англ.** *ANalysis Of VAriance*)

В случае однофакторного анализа == ***t*-критерию Стьюдента** для независимых выборок, и величина  $F$ -статистики равна квадрату соответствующей ***t*-статистики**.

Проверяем гипотезу о том, что средние выборок равны (или дисперсии равны).





## sklearn.feature\_selection : Feature Selection

The `sklearn.feature_selection` module implements feature selection algorithms, selection methods and the recursive feature elimination algorithm.

**User guide:** See the [Feature selection](#) section for further details.

<code>feature_selection.GenericUnivariateSelect</code> ([...])	Univariate feature selection
<code>feature_selection.SelectPercentile</code> ([...])	Select features according to a percentile of the highest scores.
<code>feature_selection.SelectKBest</code> ([score_func, k])	Select features according to the k highest scores.
<code>feature_selection.SelectFpr</code> ([score_func, alpha])	Filter: Select the p-values
<code>feature_selection.SelectFdr</code> ([score_func, alpha])	Filter: Select the p-values
<code>feature_selection.SelectFromModel</code> (estimator)	Meta-transformer for selecting features based on model weights.
<code>feature_selection.SelectFwe</code> ([score_func, alpha])	Filter: Select the p-values
<code>feature_selection.RFE</code> (estimator[, ...])	Feature ranking with recursive feature elimination.
<code>feature_selection.RFECV</code> (estimator[, step, ...])	Feature ranking with recursive feature elimination and cross-validation.
<code>feature_selection.VarianceThreshold</code> ([threshold])	Feature selector that removes all low-variance features.
<code>feature_selection.chi2</code> (X, y)	Compute chi-squared stats between each non-negative feature and class.
<code>feature_selection.f_classif</code> (X, y)	Compute the ANOVA F-value for the provided sample.
<code>feature_selection.f_regression</code> (X, y[, center])	Univariate linear regression tests.
<code>feature_selection.mutual_info_classif</code> (X, y)	Estimate mutual information for a discrete target variable.
<code>feature_selection.mutual_info_regression</code> (X, y)	Estimate mutual information for a continuous target variable.

Пирсон

Из любой модели

Для предобработки датасета

chi2 устойчивая статистика

Пакет "Стандарт"