

Лабораторная работа №3

Подготовка обучающей и тестовой выборки, кросс-валидация и подбор гиперпараметров на примере метода ближайших соседей.

Выполнил: Ли М.В. Группа: ИУ5-64Б

Задание:

1. Выберите набор данных (датасет) для решения задачи классификации или регрессии.
2. С использованием метода `train_test_split` разделите выборку на обучающую и тестовую.
3. Обучите модель ближайших соседей для произвольно заданного гиперпараметра `K`. Оцените качество модели с помощью подходящих для задачи метрик.
4. Произведите подбор гиперпараметра `K` с использованием `GridSearchCV` и/или `RandomizedSearchCV` и кросс-валидации, оцените качество оптимальной модели. Желательно использование нескольких стратегий кросс-валидации.
5. Сравните метрики качества исходной и оптимальной моделей.

In [1]:

```
from IPython.display import Image
import numpy as np
import pandas as pd
from sklearn.datasets import *
from typing import Dict, Tuple
from sklearn.model_selection import train_test_split
import seaborn as sns
import matplotlib.pyplot as plt
from operator import itemgetter
import matplotlib.ticker as ticker
import math
from sklearn.metrics import accuracy_score, balanced_accuracy_score
from sklearn.metrics import plot_confusion_matrix
from sklearn.metrics import precision_score, recall_score, f1_score, classification_report
from sklearn.metrics import confusion_matrix
from sklearn.metrics import mean_absolute_error, mean_squared_error, mean_squared_log_error, median_absolute_error
from sklearn.metrics import roc_curve, roc_auc_score
from sklearn.neighbors import KNeighborsRegressor, KNeighborsClassifier
from sklearn.model_selection import cross_val_score, cross_validate
from sklearn.model_selection import KFold, RepeatedKFold, LeaveOneOut, LeavePOut, ShuffleSplit, StratifiedKFold
from sklearn.preprocessing import MinMaxScaler, StandardScaler, Normalizer
from sklearn.model_selection import GridSearchCV, RandomizedSearchCV
from sklearn.model_selection import learning_curve, validation_curve
%matplotlib inline
sns.set(style="whitegrid")
```

Выборка датасета и ее разделение на тестовую и обучающую

In [2]:

```
breast_cancer = load_breast_cancer()
```

In [3]:

```
for x in breast_cancer:
    print(x)
```

```
data
target
frame
target_names
DESCR
feature_names
filename
```

In [4]:

```
# Сформируем DataFrame
breast_cancer_df = pd.DataFrame(data= np.c_[breast_cancer['data'], breast_cancer['target']],
                                columns= list(breast_cancer['feature_names']) + ['target'])
```

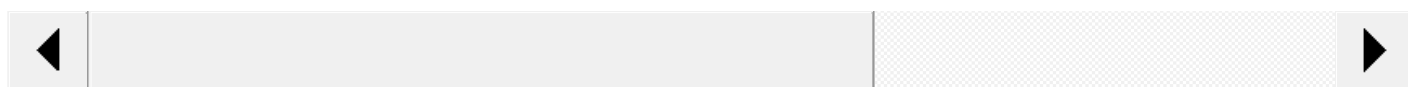
In [5]:

```
breast_cancer_df
```

Out[5]:

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mean concavity	mean concave points	mean symmetry	mean fractal dimension	...	worst texture	worst perimeter	worst area	sm
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710	0.2419	0.07871	...	17.33	184.60	2019.0	
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017	0.1812	0.05667	...	23.41	158.80	1956.0	
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790	0.2069	0.05999	...	25.53	152.50	1709.0	
3	11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520	0.2597	0.09744	...	26.50	98.87	567.7	
4	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430	0.1809	0.05883	...	16.67	152.20	1575.0	
...
564	21.56	22.39	142.00	1479.0	0.11100	0.11590	0.24390	0.13890	0.1726	0.05623	...	26.40	166.10	2027.0	
565	20.13	28.25	131.20	1261.0	0.09780	0.10340	0.14400	0.09791	0.1752	0.05533	...	38.25	155.00	1731.0	
566	16.60	28.08	108.30	858.1	0.08455	0.10230	0.09251	0.05302	0.1590	0.05648	...	34.12	126.70	1124.0	
567	20.60	29.33	140.10	1265.0	0.11780	0.27700	0.35140	0.15200	0.2397	0.07016	...	39.42	184.60	1821.0	
568	7.76	24.54	47.92	181.0	0.05263	0.04362	0.00000	0.00000	0.1587	0.05884	...	30.37	59.16	268.6	

569 rows × 31 columns



In [6]:

```
sc = MinMaxScaler()
breast_cancer_sc = sc.fit_transform(breast_cancer.data)
```

In [7]:

```
X_train, X_test, Y_train, Y_test = train_test_split(
breast_cancer_sc, breast_cancer.target, test_size=0.33, random_state=1)
```

In [8]:

```
# Размер обучающей выборки
X_train.shape, Y_train.shape
```

Out[8]:

```
((381, 30), (381,))
```

In [9]:

```
# Размер тестовой выборки
X_test.shape, Y_test.shape
```

Out[9]:

```
((188, 30), (188,))
```

In [10]:

```
X_train
```

Out[10]:

```
array([[0.55274741, 0.25059182, 0.53631401, ..., 0.57525773, 0.26197516,
        0.19362456],
       [0.2607317 , 0.24146094, 0.24462719, ..., 0.0956701 , 0.06938695,
        0.04394595],
       [0.282976 , 0.29015894, 0.27910994, ..., 0.22707904, 0.32367435,
        0.11432507],
       ...,
       [0.48364807, 0.50084545, 0.48655933, ..., 0.65257732, 0.34456929,
        0.51725043],
       [0.3336173 , 0.3902604 , 0.31787713, ..., 0.27364261, 0.13029765,
        0.13859373],
       [0.28628899, 0.29455529, 0.26826066, ..., 0.17226804, 0.08318549,
        0.043618  ]])
```

In [11]:

```
Y_train
```

Out[12]:

```
array([[0.36485399, 0.14440311, 0.37613157, ..., 0.38075601, 0.24876799,
        0.24294897],
       [0.29291495, 0.30267163, 0.29154861, ..., 0.71752577, 0.46027991,
        0.41230487],
       [0.28250272, 0.21339195, 0.27192316, ..., 0.3628866 , 0.35777646,
        0.26761118],
       ...,
       [0.60717497, 0.42069665, 0.59574321, ..., 0.82061856, 0.23713779,
        0.13846255],
       [0.49642671, 0.50625634, 0.49968903, ..., 0.59140893, 0.1172876 ,
        0.24898334],
       [0.29906763, 0.40108218, 0.28643494, ..., 0.19292096, 0.2113148 ,
        0.07569198]])
```

In [13]:

```
regl_1 = KNeighborsClassifier(n_neighbors=2)
regl_1.fit(X_train,Y_train)
targetl_1 = regl_1.predict(X_test)
len(targetl_1), targetl_1
```

Out[13]:

```
(188,
array([[1, 0, 1, 0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1,
       0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1,
       1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1,
       1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1,
       0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0,
       0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1,
       1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1]))
```

In [14]:

```
accuracy score(Y_test, target1 1)
```

Out[14]:

0.9574468085106383

In [15]:

```
def accuracy_score_for_classes(
    y_true: np.ndarray,
    y_pred: np.ndarray) -> Dict[int, float]:
    d = {'t': y_true, 'p': y_pred}
    df = pd.DataFrame(data=d)
    classes = np.unique(y_true)
    res = dict()
    for c in classes:
        temp_data_flt = df[df['t']==c]
        temp_acc = accuracy_score(
            temp_data_flt['t'].values,
            temp_data_flt['p'].values)
```

```

        res[c] = temp_acc
    return res

def print_accuracy_score_for_classes(
    y_true: np.ndarray,
    y_pred: np.ndarray):
    accs = accuracy_score_for_classes(y_true, y_pred)
    if len(accs)>0:
        print('Метка \t Accuracy')
    for i in accs:
        print('{} \t {}'.format(i, accs[i]))

```

In [16]:

```

# 2 ближайших соседа
print_accuracy_score_for_classes(Y_test, target1_1)

```

```

Метка    Accuracy
0    0.9538461538461539
1    0.959349593495935

```

Кросс-валидация

In [17]:

```

scoring = {'precision': 'precision_weighted',
           'recall': 'recall_weighted',
           'f1': 'f1_weighted'}

```

In [18]:

```

kf = KFold(n_splits=5)
scores = cross_validate(KNeighborsClassifier(n_neighbors=2),
                        breast_cancer_sc, breast_cancer.target, scoring=scoring,
                        cv=kf, return_train_score=True)

scores

```

Out[18]:

```

{'fit_time': array([0.00601196, 0.00299382, 0.00311136, 0.00397801, 0.00299215]),
 'score_time': array([0.01454091, 0.0123682 , 0.01199245, 0.0145328 , 0.00716782]),
 'test_precision': array([0.95613188, 0.95010337, 0.96514312, 0.9591089 , 0.93459708]),
 'train_precision': array([0.98734066, 0.97166449, 0.97727737, 0.97332362, 0.97722415]),
 'test_recall': array([0.95614035, 0.94736842, 0.96491228, 0.95614035, 0.92035398]),
 'train_recall': array([0.98681319, 0.96923077, 0.97582418, 0.97142857, 0.97587719]),
 'test_f1': array([0.95605944, 0.94756499, 0.96469636, 0.95682492, 0.92340952]),
 'train_f1': array([0.98688352, 0.96948639, 0.97595915, 0.97156884, 0.9759728 ])}

```

In [19]:

```

kfl = LeaveOneOut()
scores1 = cross_validate(KNeighborsClassifier(n_neighbors=2),
                        breast_cancer_sc, breast_cancer.target, scoring=scoring,
                        cv=kfl, return_train_score=True)

scores1

```

C:\Users\enjoy\anaconda3\lib\site-packages\sklearn\metrics_classification.py:1221: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

```

    _warn_prf(average, modifier, msg_start, len(result))

```

C:\Users\enjoy\anaconda3\lib\site-packages\sklearn\metrics_classification.py:1221: UndefinedMetricWarning: Recall is ill-defined and being set to 0.0 in labels with no true samples. Use `zero_division` parameter to control this behavior.

```

    _warn_prf(average, modifier, msg_start, len(result))

```

C:\Users\enjoy\anaconda3\lib\site-packages\sklearn\metrics_classification.py:1221: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

```

    _warn_prf(average, modifier, msg_start, len(result))

```

C:\Users\enjoy\anaconda3\lib\site-packages\sklearn\metrics_classification.py:1221: UndefinedMetricWarning: Recall is ill-defined and being set to 0.0 in labels with no true samples. Use `zero_division` parameter to control this behavior.

```

    _warn_prf(average, modifier, msg_start, len(result))

```

C:\Users\enjoy\anaconda3\lib\site-packages\sklearn\metrics_classification.py:1221: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

```

    _warn_prf(average, modifier, msg_start, len(result))

```

C:\Users\enjoy\anaconda3\lib\site-packages\sklearn\metrics_classification.py:1221: UndefinedMetricWarning: Recall is ill-defined and being set to 0.0 in labels with no true samples. Use `zero_division` parameter to control this behavior.

```

    _warn_prf(average, modifier, msg_start, len(result))

```

C:\Users\enjoy\anaconda3\lib\site-packages\sklearn\metrics_classification.py:1221: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

```

    _warn_prf(average, modifier, msg_start, len(result))

```

[illegible]

[illegible]

0.0029614 , 0.00299096, 0.00399232, 0.00498486, 0.00299239,
0.0039885 , 0.00598454, 0.00398898, 0.0069828 , 0.00398922,
0.00395465, 0.00399208, 0.00401902, 0.00398874, 0.00498581,
0.00399685, 0.00399542, 0. , 0. , 0. ,
0. , 0. , 0. , 0. , 0. ,
0. , 0. , 0. , 0. , 0. ,
0. , 0. , 0. , 0. , 0. ,
0.01561999, 0.0039897 , 0.01562142, 0. , 0.01562119,
0. , 0. , 0. , 0. , 0.01562142,
0. , 0. , 0.01562166, 0. , 0. ,
0. , 0. , 0. , 0. , 0.01561904,
0. , 0. , 0. , 0.01562524, 0. ,
0. , 0. , 0. , 0. , 0. ,
0. , 0. , 0. , 0.01562262, 0. ,
0.00299144, 0.00295711, 0.00299072, 0.0030241 , 0.00299144,
0.00299025, 0.00299883, 0. , 0. , 0. ,
0.0029912 , 0.00299025, 0.00299287, 0.00299668, 0.00299144,
0.00299168, 0.00299025, 0.00299168, 0.00395989, 0.00299144,
0.00399089, 0.00299168, 0.00398827, 0.00498581, 0.00398827,
0.00398946, 0.00398946, 0.00398922, 0.0040071 , 0.00299215,
0.0029912 , 0.00299144, 0.0029912 , 0.00398827, 0.00398946,
0.00299048, 0.00299263, 0.00299072, 0.00299048, 0.003021 ,
0.00401044, 0.00399089, 0.00495458, 0.00825667, 0.00897598,
0.00498676, 0.00598359, 0.0049839 , 0.00897598, 0.00849414,
0.00698018, 0.00598335, 0.00598669, 0.00498557, 0.00698066,
0.00797939, 0.00498581, 0.00598669, 0.00997543, 0.00698066,
0.00897527, 0.00698161, 0.00700927, 0.00498652, 0.00398827,
0.00395465, 0.00398993, 0.00398779, 0.00398803, 0.00501895,
0.00299096, 0.00398397, 0.00399017, 0.00299096, 0.00398827,
0.00396276, 0.00400734, 0.00498748, 0.00299478, 0.00396585,
0.00299263, 0.00299144, 0.00498605, 0.00299335, 0.00398588,
0.00299215, 0.00398827, 0.00299215, 0.00299048, 0.00498676,
0.00399041, 0.00398898, 0.00399041, 0.00498962, 0.0039885 ,
0.00398993, 0.00398755, 0.00398922, 0.00299311, 0.00398827,
0.00398827, 0.00299096, 0.004987 , 0.00398779, 0.00299382,
0.00399566, 0.00399613, 0.00302005, 0.00299168, 0.00398803,
0.00598383, 0.00398183, 0.00297904, 0.00299048, 0.00399089,
0.00398707, 0.0039885 , 0.00395918, 0.00495672, 0.00299311,
0.0039897 , 0.00398946, 0.00398874, 0.00398922, 0.00398803,
0.00598311, 0.00398898, 0.00398827, 0.00395513, 0.00299096,
0.00299239, 0.00298715, 0.0039866 , 0.00299096, 0.00398898,
0.00303054, 0.00299215, 0.00299835, 0.00303221, 0.00395751,
0.0029912 , 0.0029912 , 0.00398803, 0.00398922, 0.00299144,
0.00299382, 0.00299311, 0.00299525, 0.00299501, 0.00398898,
0.00398874, 0.00398803, 0.0029912 , 0.0029912 , 0.00398827,
0.00299096, 0.00398898, 0.00398827, 0.00398898, 0.00398922,
0.0029912 , 0.00395751, 0.00403094, 0.00298977, 0.00398946,
0.00402021, 0.00398827, 0.00303245, 0.00299215, 0.00299215,
0.0039897 , 0.00402069, 0.00294852, 0.00498843, 0.00403428,
0.00398874, 0.00398803, 0.00299168, 0.00498796, 0.00299883,
0.00398874, 0.0029912 , 0.00398922, 0.00398946, 0.00299168,
0.00398898, 0.00299191, 0.00299001, 0.00398922, 0.00402236,
0.00299263, 0.00299239, 0.00295806, 0.00302672, 0.00299072,
0.00403762, 0.00399017, 0.00299168, 0.00298905, 0.00299501,
0.00302315, 0.00300241, 0.0040071 , 0.00299144, 0.00399041,
0.00403118, 0.0039866 , 0.00399089, 0.00498915, 0.00299454,
0.00299263, 0.00396371, 0.00298929, 0.00298858, 0.00302601,
0.00395966, 0.00398827, 0.00299144, 0.00299144, 0.0039916 ,
0.00296235, 0.00298929, 0.00299215, 0.00299549, 0.00296187,
0.00299215, 0.00395942, 0.00299144, 0.00299191, 0.00299191,
0.00299454, 0.00298929, 0.00303054, 0.00299048, 0.00299191,
0.00299191, 0.00296164, 0.00299144, 0.00299215, 0.00299072,
0.00498581, 0.00299168, 0.0029912 , 0.00299191, 0.00298977,
0.00299096, 0.00299096, 0.0029912 , 0.00299168, 0.00299144,
0.00299144, 0.0029912 , 0.00299168, 0.00299191, 0.00299144,
0.00299144, 0.00299191, 0.00299263, 0.0039916 , 0.00298953,
0.00299144, 0.00299215, 0.00399208, 0.00398779, 0.00398803,
0.00299168, 0.00398874, 0.00299168, 0.00299191, 0.0039897 ,
0.00399089, 0.00299191, 0.00399089, 0.00398874, 0.0039897 ,
0.00299239, 0.00299168, 0.00299215, 0.00398803, 0.00398946,
0.00299191, 0.00299191, 0.00295758, 0.00302434, 0.00299191,
0.00299168, 0.00299168, 0.00299096, 0.00299168, 0.002949 ,
0.00299191, 0.00299144, 0.00299215, 0.00299239, 0.00295544,
0.0029881 , 0.00302458, 0.00302029, 0.00303149, 0.00302362,
0.00299215, 0.00299096, 0.00296235, 0.00299048, 0.00299215,
0.00299239, 0.00398993, 0.00301862, 0.00299263, 0.00299191,
0.00299215, 0.00299191, 0.00299191, 0.00299191, 0.00299191,


```
0.00302124, 0.00402141, 0.00299191, 0.00299478, 0.00299215,
0.00299191, 0.00299191, 0.0029912 , 0.00299191, 0.00299072,
0.00299168, 0.00299263, 0.00299239, 0.00299191, 0.00299263,
0.00299311, 0.00301981, 0.00299454, 0.00299096, 0.00299072,
0.00299144, 0.00301981, 0.00299191, 0.00298762, 0.00299191,
0.00299191, 0.00199413, 0.00303054, 0.00299263, 0.00299215,
0.00299191, 0.00299215, 0.00299215, 0.00299215, 0.00299215,
0.00299168, 0.00299191, 0.00299215, 0.00299215, 0.00195527,
0.00298882, 0.00299191, 0.0029912 , 0.00299215, 0.00299191,
0.00299096, 0.00299144, 0.00299191, 0.00399041, 0.0029521 ,
0.00296092, 0.00299239, 0.00299215, 0.00299096, 0.00299215,
0.00299168, 0.00398922, 0.00395322, 0.00399041]),
'score_time': array([0.00398636, 0. , 0.00501871, 0.0039928 , 0.00396776,
0. , 0. , 0. , 0.00296092, 0.00405478,
0.00398922, 0.00398564, 0.00363636, 0.00399685, 0.00301909,
0.00295997, 0.00398946, 0.00298905, 0.00398922, 0.00299168,
0.00299191, 0.00298667, 0.00299239, 0. , 0. , 0. ,
0.01565051, 0. , 0. , 0. , 0.00299239, 0.0029912 ,
0.00298882, 0.01008797, 0.0039854 , 0.00299215, 0.00399113,
0. , 0.00399256, 0.00399065, 0.00299191, 0.00338149,
0.00698066, 0.00399017, 0.00299311, 0.00299311, 0.00598359,
0.0029974 , 0.00399971, 0.00299239, 0.00298715, 0.00797915,
0.0039885 , 0.00398922, 0.00498676, 0.0059886 , 0.00498462,
0.0059855 , 0.00598431, 0.00498271, 0.00798082, 0.00498772,
0.00398922, 0.00399303, 0. , 0. , 0. , 0. ,
0.0156219 , 0.00134587, 0. , 0.00395274, 0.00399041,
0.0156219 , 0.01562309, 0. , 0.00495028, 0.00498581,
0.00498343, 0.00402761, 0.00399375, 0.00313234, 0. , 0. ,
0.01562667, 0.01559162, 0. , 0.00401592, 0.0029664 ,
0. , 0.00296736, 0.00304103, 0.00299311, 0.00399017,
0.00301743, 0. , 0.00302148, 0.00398684, 0.00398803,
0.00299001, 0. , 0. , 0.00299239, 0.00243783,
0. , 0.00402617, 0.01565099, 0.00298905, 0. , 0. ,
0.00299168, 0.00299311, 0. , 0.01562381, 0.00396156,
0. , 0. , 0.00299263, 0.00398946, 0.00401998,
0. , 0.00399423, 0. , 0.00596762, 0.00398922,
0.00399208, 0.00498056, 0.00399041, 0.00398874, 0.00299072,
0.0039947 , 0.00601196, 0.00398612, 0.00598764, 0.00398684,
0.00498986, 0.00398946, 0.00398707, 0.0039897 , 0.00399995,
0.00399065, 0.00498533, 0.00395942, 0.00299263, 0.00299239,
0.00298691, 0.00297928, 0. , 0. , 0.01559091,
0.01562405, 0.01562905, 0.01561666, 0.01563239, 0.01562452,
0. , 0.01562452, 0.01561832, 0. , 0. , 0. ,
0. , 0. , 0.01562738, 0. , 0. , 0. ,
0. , 0.00144982, 0. , 0. , 0. , 0. ,
0. , 0. , 0.01562095, 0.01562047, 0. , 0. ,
0.0156219 , 0.01562119, 0. , 0.01562238, 0.01562214,
0. , 0. , 0. , 0. , 0. , 0. ,
0.01561975, 0. , 0. , 0. , 0. , 0.01562047,
0. , 0. , 0.01562214, 0. , 0. , 0. ,
0.01562285, 0. , 0. , 0. , 0. , 0. ,
0.00199461, 0.00299191, 0.00296545, 0.00299144, 0.00199461,
0.00299263, 0.00300217, 0. , 0. , 0.00662112,
0.00296283, 0.00301957, 0.00299168, 0.00495434, 0.00302291,
0.00299239, 0.00402141, 0.00301957, 0.00401449, 0.00299239,
0.00501251, 0.00299191, 0.00399137, 0.00498581, 0.00398946,
0.00399184, 0.00299191, 0.00299406, 0.00299263, 0.00299191,
0.00299287, 0.00299287, 0.00398993, 0.00299335, 0.00299311,
0.00299263, 0.00299144, 0.00398803, 0.00299191, 0.00298929,
0.00296998, 0.00299263, 0.00997663, 0.00821519, 0.00997305,
0.00698304, 0.00897574, 0.00698423, 0.00897646, 0.00897861,
0.01097083, 0.00498509, 0.00398874, 0.00399065, 0.00797939,
0.00698066, 0.00698209, 0.00698113, 0.02493095, 0.00698304,
0.00897646, 0.00797701, 0.00698209, 0.00598621, 0.00299263,
0.00402164, 0.00398898, 0.00399113, 0.00398993, 0.00298834,
0.00399232, 0.00299215, 0.00299311, 0.00399065, 0.00299215,
0.0029912 , 0.00300145, 0.00498676, 0.00299239, 0.00301123,
0.00296092, 0.00299239, 0.00498557, 0.00398874, 0.00398803,
0.00498772, 0.00299287, 0.0039897 , 0.00399113, 0.00598431,
0.00299191, 0.00398946, 0.00299096, 0.00398612, 0.00398993,
0.0029912 , 0.00299239, 0.00398827, 0.00399065, 0.00299287,
0.00299358, 0.00399113, 0.00398779, 0.00299406, 0.00402594,
0.00299072, 0.00298667, 0.00398993, 0.00299287, 0.00298643,
0.00698352, 0.00299501, 0.00399089, 0.00399017, 0.00299335,
0.00299263, 0.00299311, 0.00402498, 0.00501633, 0.00399017,
0.00298953, 0.00299215, 0.00398755, 0.00299406, 0.00299191,
0.00299168, 0.00299215, 0.00299382, 0.00299263, 0.00403452,
```

0.00302911, 0.00399637, 0.00299764, 0.00299168, 0.00299239,
0.00298667, 0.00299239, 0.00298691, 0.00298715, 0.00299168,
0.00299335, 0.00299287, 0.00299287, 0.00299239, 0.00299358,
0.00398874, 0.00395799, 0.00295663, 0.00299001, 0.00299263,
0.00299072, 0.00299287, 0.00299287, 0.00299287, 0.00299311,
0.00498533, 0.00299001, 0.00299215, 0.00299239, 0.00199533,
0.00299287, 0.00298929, 0.00295353, 0.00299191, 0.00299287,
0.00295854, 0.00299335, 0.00295186, 0.00398803, 0.00299525,
0.00299335, 0.00295901, 0.00402188, 0.00299001, 0.002949 ,
0.00299215, 0.00299239, 0.00299191, 0.00398803, 0.002985 ,
0.00299239, 0.00398874, 0.00299215, 0.00299215, 0.00299168,
0.0029912 , 0.00299215, 0.00399113, 0.00299239, 0.00299168,
0.00301886, 0.00298166, 0.00299311, 0.00299287, 0.00299358,
0.00298333, 0.00298977, 0.00398898, 0.0039897 , 0.00398779,
0.00299215, 0.00394702, 0.00300479, 0.00498772, 0.00302887,
0.00295305, 0.00399184, 0.00299311, 0.00299168, 0.00299072,
0.00299335, 0.00302315, 0.00299191, 0.00299025, 0.00298977,
0.00399041, 0.00299382, 0.00299263, 0.00301981, 0.00395465,
0.00299215, 0.00299215, 0.00303125, 0.00198865, 0.00302434,
0.00299168, 0.00298977, 0.00299215, 0.00299215, 0.00299215,
0.00298977, 0.00299168, 0.00295496, 0.00299263, 0.00299215,
0.00299168, 0.00299263, 0.00299287, 0.00199509, 0.00199556,
0.00398946, 0.00199533, 0.00299263, 0.00299335, 0.0029943 ,
0.00299287, 0.0029912 , 0.0019958 , 0.00299239, 0.00299239,
0.00199533, 0.00299239, 0.00299239, 0.00398993, 0.00299215,
0.00299025, 0.00398922, 0.00298953, 0.00299215, 0.00299144,
0.00299239, 0.00299144, 0.00398636, 0.00299215, 0.0029943 ,
0.00299454, 0.00299215, 0.00299263, 0.0039897 , 0.0029912 ,
0.00299144, 0.00398993, 0.00299144, 0.00398946, 0.00299168,
0.00299168, 0.00299883, 0.00501895, 0.00299215, 0.00299215,
0.00299239, 0.00295782, 0.00299215, 0.00299954, 0.00299239,
0.00299239, 0.00299239, 0.00299191, 0.00299239, 0.00299191,
0.00299358, 0.00299239, 0.00299168, 0.00299144, 0.00302386,
0.00200534, 0.0019927 , 0.00299096, 0.00295472, 0.00200629,
0.00299191, 0.00299025, 0.0030272 , 0.00299215, 0.00299191,
0.004987 , 0.00299144, 0.00299191, 0.00299168, 0.00299454,
0.00299215, 0.00199461, 0.00199628, 0.00299191, 0.00299287,
0.00300264, 0.00199318, 0.00299215, 0.00299168, 0.00399065,
0.00299215, 0.00299191, 0.00299215, 0.00299215, 0.00299191,
0.00299144, 0.00302148, 0.00299144, 0.00299215, 0.00199413,
0.00298905, 0.00296426, 0.00298953, 0.0029912 , 0.00299191,
0.00299335, 0.00299788, 0.00302434, 0.00295758, 0.00202751,
0.00299215, 0.00299239, 0.00295568, 0.00199461, 0.00299191,
0.00199461, 0.00299191, 0.00299191, 0.00199461, 0.00299191,
0.00299215, 0.00299239, 0.00203156, 0.00299191, 0.00299001,
0.00299168, 0.00199485, 0.00299191, 0.00199676, 0.00199485,
0.00299239, 0.00299668, 0.00302911, 0.00301957, 0.00299168,
0.00398827, 0.00299215, 0.00299191, 0.00299311, 0.00299215,
0.00299191, 0.00299215, 0.00199485, 0.00299144]],

'test_precision': array([1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
1., 1., 1., 1., 1., 0., 1., 1., 1., 1., 1., 1., 1., 1., 1., 0., 1.,
1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
1., 1., 1., 1., 0., 1., 1., 1., 1., 1., 1., 1., 1., 0., 1., 1., 1.,
1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 0.,
0., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
1., 1., 1., 0., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
1., 1., 1., 0., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
1., 1., 1., 1., 1., 1., 1., 0., 1., 1., 1., 1., 1., 1., 1., 1.,
1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
0., 1., 1., 1., 1., 1., 1., 1., 0., 1., 1., 1., 1., 1., 1., 1.,
1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
0., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
1., 1., 1., 1., 1., 1., 0., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
1., 1., 0., 1., 1., 1., 0., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
1., 1., 1., 1., 1., 1., 0., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]


```
0.97375855, 0.97375855, 0.97376143, 0.97375855, 0.97375855,
0.97375855, 0.97375855, 0.97375855, 0.97201937, 0.97375855,
0.97375855, 0.97375855, 0.97375855, 0.97375855, 0.97375855,
0.97376143, 0.97375855, 0.97376143, 0.97376143, 0.97375855,
0.97376143, 0.97375855, 0.97375855, 0.97375855, 0.97375855,
0.97375855, 0.97376143, 0.97375855, 0.97375855, 0.97376143,
0.97375855, 0.97376143, 0.97375855, 0.97549884, 0.97376143,
0.97375855, 0.97376143, 0.97375855, 0.97375855, 0.97375855,
0.97375855, 0.97375855, 0.97375855, 0.97375855, 0.97375855,
0.97376143, 0.97376143, 0.97375855, 0.97375855, 0.97375855,
0.97375855, 0.97201937, 0.97375855, 0.97376143, 0.97375855,
0.97375855, 0.97375855, 0.97375855, 0.97375855, 0.97375855,
0.97375855, 0.97375855, 0.97375855, 0.97375855, 0.97376143,
0.97375855, 0.97375855, 0.97375855, 0.97375855, 0.97376143,
0.97375855, 0.97375855, 0.97376143, 0.97375855, 0.97375855,
0.97375855, 0.97549884, 0.97375855, 0.97376143, 0.97376143,
0.97549884, 0.97376143, 0.97375855, 0.97376143, 0.97375855,
0.97375855, 0.97375855, 0.97375855, 0.97375855, 0.97376143,
0.97375855, 0.97375855, 0.97375855, 0.97375855, 0.97376143,
0.97375855, 0.97375855, 0.97376143, 0.97375855, 0.97724245,
0.97375855, 0.97376143, 0.97376143, 0.97375855, 0.97375855,
0.97375855, 0.97376143, 0.97375855, 0.97375855, 0.97375855,
0.97375855, 0.97549884, 0.97375855, 0.97375855, 0.97375855,
0.97375855, 0.97375855, 0.97375855, 0.97376143, 0.97375855,
0.97376143, 0.97376143, 0.97375855, 0.97375855, 0.97375855,
0.97375855, 0.97549884, 0.97201937, 0.97375855, 0.97375855,
0.97375855, 0.97375855, 0.97375855, 0.97375855, 0.97375855,
0.97375855, 0.97375855, 0.97375855, 0.97375855, 0.97375855,
0.97375855, 0.97375855, 0.97375855, 0.97375855, 0.97375855,
0.97549884, 0.97375855, 0.97376143, 0.97376143, 0.97376143,
0.97376143, 0.97376143, 0.97376143, 0.97375855]]})
```

Оптимизация гиперпараметров

In [20]:

```
n_range = np.array(range(1,70,2))
tuned_parameters = [{'n_neighbors': n_range}]
tuned_parameters
```

Out[20]:

```
[{'n_neighbors': array([ 1,  3,  5,  7,  9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33,
                        35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67,
                        69])}]
```

In [21]:

```
#GridSearch
clf_gs = GridSearchCV(KNeighborsClassifier(), tuned_parameters, cv=kf1, scoring='accuracy')
clf_gs.fit(X_train, Y_train)
```

Out[21]:

```
GridSearchCV(cv=LeaveOneOut(), estimator=KNeighborsClassifier(),
             param_grid=[{'n_neighbors': array([ 1,  3,  5,  7,  9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31,
                        33,
                        35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67,
                        69])}],
             scoring='accuracy')
```

In [22]:

```
clf_gs.cv_results_
```

Out[22]:

```
{'mean_fit_time': array([0.0018117, 0.00174626, 0.00199766, 0.00215117, 0.00241277,
                        0.00246673, 0.00208115, 0.00353876, 0.00243144, 0.00281325,
                        0.00261019, 0.00202312, 0.0023389, 0.00200821, 0.00172697,
                        0.00193674, 0.00192714, 0.00163552, 0.00190436, 0.00240536,
                        0.00243987, 0.00209576, 0.00216749, 0.0023078, 0.0019885,
                        0.00225676, 0.00217317, 0.00139405, 0.00230241, 0.00180394,
                        0.00218051, 0.0025425, 0.00233556, 0.0017407, 0.00225984]),
 'std_fit_time': array([0.00046011, 0.00050877, 0.00047894, 0.00053819, 0.00066837,
                        0.00259145, 0.00255815, 0.00148311, 0.00445789, 0.00535073,
                        0.00529592, 0.00517541, 0.004428, 0.00366948, 0.00398377,
                        0.00501181, 0.00513734, 0.00473014, 0.00477785, 0.0032469,
                        0.00449573, 0.00438595, 0.00535557, 0.00443389, 0.00479774,
                        0.00544717, 0.00540622, 0.00445354, 0.00324098, 0.0048279,
                        0.00385968, 0.00465127, 0.00552328, 0.00424601, 0.00294394]),
 'mean_score_time': array([0.00084551, 0.00082947, 0.00101601, 0.00108539, 0.00140011,
                        0.00142054, 0.00110113, 0.00189557, 0.00152325, 0.00080137,
                        0.00094293, 0.00093373, 0.00124703, 0.00097939, 0.00098791,
                        0.00087333, 0.00081994, 0.00116769, 0.00108412, 0.00134131,
```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]


```
{'n_neighbors': 15}
```

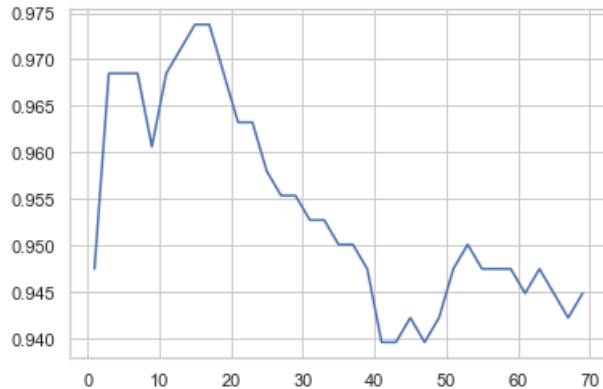
Out[25]:

```
# Изменение качества на тестовой выборке в зависимости от K-соседей
plt.plot(n_range, clf_gs.cv_results_['mean_test_score'])
```

In [26]:

```
[<matplotlib.lines.Line2D at 0x2020d3c2970>]
```

Out[26]:



In [27]:

```
#RandomizedSearchCV
clf_rs = RandomizedSearchCV(KNeighborsClassifier(), tuned_parameters, cv=kf1, scoring='accuracy')
clf_rs.fit(X_train, Y_train)
```

Out[27]:

```
RandomizedSearchCV(cv=LeaveOneOut(), estimator=KNeighborsClassifier(),
                    param_distributions=[{'n_neighbors': array([ 1,  3,  5,  7,  9, 11, 13, 15, 17, 19, 21, 23,
25, 27, 29, 31, 33,
35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67,
69])}],
                    scoring='accuracy')
```

In [28]:

```
# Лучшее значение параметров
clf_rs.best_params_
```

Out[28]:

```
{'n_neighbors': 15}
```

In [29]:

```
# Лучшая модель
clf_rs.best_estimator_
```

Out[29]:

```
KNeighborsClassifier(n_neighbors=15)
```

K-fold

In [30]:

```
clf_gs = GridSearchCV(KNeighborsClassifier(), tuned_parameters, cv=kf, scoring='accuracy')
clf_gs.fit(X_train, Y_train)
```

Out[30]:

```
GridSearchCV(cv=KFold(n_splits=5, random_state=None, shuffle=False),
              estimator=KNeighborsClassifier(),
              param_grid=[{'n_neighbors': array([ 1,  3,  5,  7,  9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31,
33,
35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67,
69])}],
              scoring='accuracy')
```

In [31]:

```
clf_gs.best_score_, clf_gs.best_params_
```

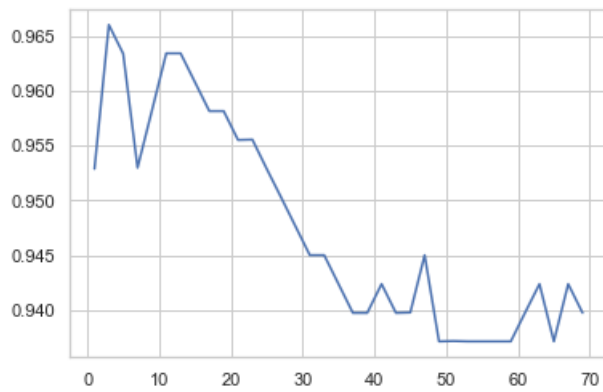
Out[31]:

```
(0.9659945317840055, {'n_neighbors': 3})
```

In [32]:

```
# Изменение качества на тестовой выборке в зависимости от K-соседей
plt.plot(n_range, clf_gs.cv_results_['mean_test_score'])
```

[<matplotlib.lines.Line2D at 0x2020d23da00>]



Out[32]:



In [33]:

```
clf_rs = RandomizedSearchCV(KNeighborsClassifier(), tuned_parameters, cv=kf, scoring='accuracy')
clf_rs.fit(X_train, Y_train)
```

Out[33]:

```
RandomizedSearchCV(cv=KFold(n_splits=5, random_state=None, shuffle=False),
                    estimator=KNeighborsClassifier(),
                    param_distributions=[{'n_neighbors': array([ 1,  3,  5,  7,  9, 11, 13, 15, 17, 19, 21, 23,
25, 27, 29, 31, 33,
35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67,
69])}],
                    scoring='accuracy')
```

In [34]:

```
clf_rs.best_score_, clf_rs.best_params_
```

Out[34]:

```
(0.9659945317840055, {'n_neighbors': 3})
```

In []: