```
Линейные модели, SVM и деревья решений
In [1]:
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import SGDClassifier
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
from sklearn.metrics import precision_score, recall_score
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import GridSearchCV
%matplotlib inline
sns.set(style="ticks")
In [2]:
data=pd.read_csv('heart.csv', sep=",")
In [3]:
data.dtypes
Out[3]:
        int64
age
sex
        int64
ср
        int64
trestbps
        int64
chol
        int64
         int64
restecg
         int64
thalach
slope
         int64
        int64
ca
thal
        int64
target
        int64
dtype: object
In [4]:
data.isnull().sum()
Out[4]:
age
        0
        0
sex
        0
ср
trestbps
         0
chol
        0
restecg
         0
thalach
         0
         0
slope
ca
        0
        0
thal
target
dtype: int64
In [5]:
data.head()
Out[5]:
             cp trestbps chol restecg thalach slope ca
                                                           thal target
     63
              3
                      145
                           233
                                            150
     37
           1
              2
                      130
                           250
                                            187
                                                         0
```

age sex cp	trestbps	chol res	stecg th	halach 178	slope 2	0	2	1	_	_	_	<u>.</u>	i-	Ĺ.	1	1	\rightarrow	get 1	4	-	—	4	1	4	1	4	1—	4	1	1	4	•		
4 57 0 0	120	354	1	163	2	0	2	1				1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Разделение вы	борки на с	обучающу	ю и тес	TORVIO																														
	oopkii na c	осу ісіощу	10 11 100	or obylo																														
In [6]: X_train, X_test,	y train, y	test = trair	n test s	split(
data, data['tar					= 1)																													
In [7]:																	_	_	_												_			
# Размер обуча X_train.shape, у																																		
Out[7]:																																		
((242, 11), (242,))																																	
In [8]:																																		
# Размер тесто X_test.shape, у_																																		
Out[8]:																																		
((61, 11), (61,))																																		
In [9]:																																		
#обучение лине sgd = SGDClass			roin)																			_												
		,_	rairi)																															
en added in SGI	OClassifier lefault max	3\lib\site-p in 0.19. If	ackage both ar	e left ur	nset, th	ey de	fault to	o max	astic ax_i	stic	astidax_	asti	asti ıax_	ast ıax_	ıast ıax _.	nax	max	max	max	nax	nax	max	nax	max	max	nax	ıası ıax	max	nax	nax	max	nax	max	max
en added in SGI 00. From 0.21, c FutureWarning	OClassifier lefault max	3\lib\site-p in 0.19. If	ackage both ar	e left ur	nset, th	ey de	fault to	o max	astic ax_i	stic	astidax_	asti ıax_	asti nax_	ast lax_	ıast ıax	nax	max	max	max	nax	nax	max	nax	max	max	nax	ıası	max	nax	nas	max	nax	max	max
en added in SGI 00. From 0.21, c FutureWarning n [10]:	OClassifier lefault max)	3∖lib∖site-p in 0.19. If :_iter will b	ackage both ar	e left ur	nset, th	ey de	fault to	o max	astic ax_i	stic	astidax_	asti nax_	asti nax_	ast ıax_	ıast ıax _.	nax	max	max	max	nax	nax	max	nax	max	max	nax	nax	max	nax	nax	max	nax	max	max
en added in SGI 00. From 0.21, c FutureWarning In [10]: target_sgd = sgo	OClassifier lefault max)	3∖lib∖site-p in 0.19. If :_iter will b	ackage both ar	e left ur	nset, th	ey de	fault to	o max	astic ax_i	stic	astid	asti nax_	asti	ast lax_	ıast ıax	nax	max	max	max	nax	nax	max	nax	max	max	nax	ıas	max	nax	nax	max	nax	max	max
C:\Users\Dovlat\ en added in SGI 00. From 0.21, c FutureWarning In [10]: target_sgd = sgc In [11]: accuracy_score precision_score recall_score(y_t	OClassifier lefault max) d.predict(X (y_test, tar (y_test, tar (y_test, tar	3\lib\site-p in 0.19. If t_iter will b test) get_sgd), get_sgd), get_sgd),	both ar	e left ur	nset, th	ey de	fault to	o max	astic	stic	astid	asti nax_	asti	ast lax_	nast	nax	max	max	max	nax	nax	max	nax	max	max	nax	nas	max	nax	nax	max	nax	max	max
en added in SGI DO. From 0.21, of FutureWarning In [10]: Earget_sgd = sgd In [11]: Eaccuracy_score Drecision_score Precall_score(y_t	OClassifier lefault max) d.predict(X (y_test, tar (y_test, tar (y_test, tar	3\lib\site-p in 0.19. If t_iter will b test) get_sgd), get_sgd), get_sgd),	both ar	e left ur	nset, th	ey de	fault to	o max	astic ax_i	estic	astid	asti	asti	ast ıax_	nast	nax	max	max	max	nax	nax	max	nax	max	max	nax	ıas ıax	max	nax	nax	max	nax	max	max
en added in SGI 00. From 0.21, o FutureWarning In [10]: target_sgd = sgd In [11]: accuracy_score precision_score recall_score(y_t Out[11]:	OClassifier lefault max) d.predict(X (y_test, tar (y_test, tar est, target_	3\lib\site-p in 0.19. If t_iter will b t_test) get_sgd), get_sgd), sgd)	both ar be 1000,	re left ui	nset, th	ey de	fault to	o max	astic	stic	astid	asti	asti	ast lax_	nast nax	nax	max) max	max	nax	nax	max	nax	max	max	nax	nax	max	nax	nax	max	nax	max	max
en added in SGI 20. From 0.21, o FutureWarning In [10]: Earget_sgd = sgd In [11]: Eaccuracy_score Orecision_score Orecall_score(y_t C0.70491803278 In [12]:	OClassifier lefault max) d.predict(X (y_test, tar (y_test, tar est, target_	3\lib\site-p in 0.19. If t_iter will b t_test) get_sgd), get_sgd), sgd)	both ar be 1000,	re left ui	nset, th	ey de	fault to	o max	astic	stic	astid	asti lax_	asti	ast lax_	ıast ıax <u>.</u>	nax	max) max	max	nax	nax	max	nax	max	max	nax	nas	max	nax	nax	max	nax	max	max
en added in SGI 20. From 0.21, o FutureWarning In [10]: Earget_sgd = sgd In [11]: Eaccuracy_score Precision_score Precision_score(y_t In [11]: In [11]: In [12]: In [OClassifier lefault max d.predict(X (y_test, tar (y_test, tar est, target_	3\lib\site-p in 0.19. If c_iter will b c_test) get_sgd), get_sgd), sgd)	package both ar pe 1000,	re left ui, and de	nset, th	ey de	fault to	o max	astic	stic	astic	asti	asti	ast lax_	nast	nax	max) max	max	nax	nax	max	nax	max	max	nax	nas	max	nax	nax	max	nax	max	max
en added in SGI 10. From 0.21, o FutureWarning In [10]: In [11]: In [11]: In [11]: In [11]: In [11]: In [11]: In [12]: In [12]: In [13]: In [13]:	OClassifier lefault max) d.predict(X (y_test, tar (y_test, tar est, target_ 68853, 0.7	3\lib\site-p in 0.19. If t_iter will b test) get_sgd), get_sgd), sgd) 709677419	package both ar pe 1000,	re left ui, and de	nset, th	ey de	fault to	o max	astic	stic	astid	asti nax_	asti	ast lax_	ıast ıax	nax	max) max	max	nax	nax	max	nax	max	max	nax	nas	max	nax	nax	max	nax	max	max
en added in SGI 20. From 0.21, o FutureWarning In [10]: In [11]: In [12]: In [12]: In [13]: In [13]: In [13]:	OClassifier lefault max) d.predict(X (y_test, tar (y_test, tar est, target_ 68853, 0.7	3\lib\site-p in 0.19. If t_iter will b test) get_sgd), get_sgd), sgd) 709677419	package both ar pe 1000,	re left ui, and de	nset, th	ey de	fault to	o max	astic	stic	astic	asti nax_	asti	ast lax_	nast	nax	max) max	max	nax	nax	max	nax	max	max	nax	nas	max	nax	nax	max	nax	max	max
en added in SGI 00. From 0.21, o FutureWarning In [10]: target_sgd = sgd In [11]: accuracy_score precision_score recall_score(y_t Out[11]: (0.70491803278 In [12]: #обучение SVC svc = SVC(gamil In [13]: target_svc = svc In [14]:	OClassifier lefault max d.predict(X (y_test, tar (y_test, tar est, target_ 68853, 0.7	3\lib\site-p in 0.19. If c_iter will b c_test) get_sgd), get_sgd), get_sgd) 709677419	package both ar pe 1000,	re left ui, and de	nset, th	ey de	fault to	o max	astic	stic	astid	asti	astinax_	ast lax_	nast	nax	max) max	max	nax	nax	max	nax	max	max	nax	nax	max	nax	nax	max	nax	max	max
en added in SGI 00. From 0.21, o FutureWarning In [10]: target_sgd = sgd In [11]: accuracy_score precision_score recall_score(y_t Out[11]: (0.70491803278 In [12]: #0бучение SVC svc = SVC(game In [13]: target_svc = svc	Classifier lefault max d.predict(X (y_test, tar (y_test, target_ 68853, 0.7 c.predict(X_ (y_test, target_ c.predict(X_ (y_t	3\lib\site-p in 0.19. If c_iter will b c_test) get_sgd), get_sgd), sgd) fit(X_train _test) get_svc), get_svc), get_svc),	package both ar pe 1000,	re left ui, and de	nset, th	ey de	fault to	o max	astic ax_i	stic	astic	asti nax_	asti	ast lax_	nast	nax	max) max	max	nax	nax	max	nax	max	max	nax	nas	max	nax	nas	max	nax	max	max

(0.5081967213114754, 0.5081967213114754, 1.0)In [15]: #обучение дерева решений tree = DecisionTreeClassifier(random_state=1, max_depth=0.75).fit(X_train, y_train) In [16]: target_tree = tree.predict(X_test) In [17]: accuracy_score(y_test, target_tree), \ precision_score(y_test, target_tree), \ recall_score(y_test, target_tree) Out[17]: (0.5081967213114754, 0.5081967213114754, 1.0)Подбор гиперпараметра In [18]: scores_sgd = cross_val_score(SGDClassifier(), X_train, y_train, cv=2) scores_sgd C:\Users\Dovlat\Anaconda3\lib\site-packages\sklearn\linear model\stochastic gradient.py:166: FutureWarning: max_iter and tol parameters have be en added in SGDClassifier in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=10 00. From 0.21, default max_iter will be 1000, and default tol will be 1e-3. FutureWarning) C:\Users\Dovlat\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:166: FutureWarning: max_iter and tol parameters have be en added in SGDClassifier in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=10 00. From 0.21, default max_iter will be 1000, and default tol will be 1e-3. FutureWarning) Out[18]: array([0.63636364, 0.70247934]) In [19]: scores_svm_svc = cross_val_score(SVC(gamma='auto'), X_train, y_train, cv=2) scores_svm_svc Out[19]: array([0.55371901, 0.55371901]) In [20]: scores_decision_tree = cross_val_score(DecisionTreeClassifier(), X_train, y_train, cv=2) scores_decision_tree Out[20]: array([1., 1.]) In [21]: parameters = {'alpha':[0.5,0.4,0.3,0.2,0.1]} clf_gs_sgd = GridSearchCV(SGDClassifier(), parameters, cv=2, scoring='accuracy') clf_gs_sgd.fit(X_train, y_train)

FutureWarning)

C:\Llsers\Dovlat\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.pv:166: FutureWarning: max_iter and tol parameters have be

C:\Users\Dovlat\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:166: FutureWarning: max_iter and tol parameters have be en added in SGDClassifier in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=10

00. From 0.21, default max_iter will be 1000, and default tol will be 1e-3.

en added in SGDClassifier in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=10 00. From 0.21, default max_iter will be 1000, and default tol will be 1e-3.

FutureWarning)

C:\Users\Dovlat\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:166: FutureWarning: max_iter and tol parameters have be en added in SGDClassifier in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=10 00. From 0.21, default max_iter will be 1000, and default tol will be 1e-3.

FutureWarning)

C:\Users\Dovlat\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:166: FutureWarning: max_iter and tol parameters have be en added in SGDClassifier in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=10 00. From 0.21, default max_iter will be 1000, and default tol will be 1e-3.

FutureWarning)

C:\Users\Dovlat\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:166: FutureWarning: max_iter and tol parameters have be en added in SGDClassifier in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=10 00. From 0.21, default max_iter will be 1000, and default tol will be 1e-3.

FutureWarning

C:\Users\Dovlat\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:166: FutureWarning: max_iter and tol parameters have be en added in SGDClassifier in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=10 00. From 0.21, default max_iter will be 1000, and default tol will be 1e-3.

FutureWarning)

C:\Users\Dovlat\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:166: FutureWarning: max_iter and tol parameters have be en added in SGDClassifier in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=10 00. From 0.21, default max_iter will be 1000, and default tol will be 1e-3.

FutureWarning)

C:\Users\Dovlat\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:166: FutureWarning: max_iter and tol parameters have be en added in SGDClassifier in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=10 00. From 0.21, default max_iter will be 1000, and default tol will be 1e-3.

FutureWarning)

C:\Users\Dovlat\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:166: FutureWarning: max_iter and tol parameters have be en added in SGDClassifier in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=10 00. From 0.21, default max_iter will be 1000, and default tol will be 1e-3.

FutureWarning)

C:\Users\Dovlat\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:166: FutureWarning: max_iter and tol parameters have be en added in SGDClassifier in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=10 00. From 0.21, default max_iter will be 1000, and default tol will be 1e-3.

FutureWarning)

C:\Users\Dovlat\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:166: FutureWarning: max_iter and tol parameters have be en added in SGDClassifier in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=10 00. From 0.21, default max_iter will be 1000, and default tol will be 1e-3. FutureWarning)

Out[21]:

```
GridSearchCV(cv=2, error_score='raise-deprecating', estimator=SGDClassifier(alpha=0.0001, average=False, class_weight=None, early_stopping=False, epsilon=0.1, eta0=0.0, fit_intercept=True, l1_ratio=0.15, learning_rate='optimal', loss='hinge', max_iter=None, n_iter=None, n_iter_no_change=5, n_jobs=None, penalty='l2', power_t=0.5, random_state=None, shuffle=True, tol=None, validation_fraction=0.1, verbose=0, warm_start=False), fit_params=None, iid='warn', n_jobs=None, param_grid={'alpha': [0.5, 0.4, 0.3, 0.2, 0.1]}, pre_dispatch='2*n_jobs', refit=True, return_train_score='warn', scoring='accuracy', verbose=0)
```

In [22]:

```
#для линейной модели
clf_gs_sgd.best_params_
```

▼

Out[22]:

{'alpha': 0.3}

In [23]:

```
parameters = {'gamma':[0.9,0.8,0.7,0.6,0.5,0.4,0.3,0.2,0.1]}
clf_gs_svm_svc = GridSearchCV(SVC(), parameters, cv=2, scoring='accuracy')
clf_gs_svm_svc.fit(X_train, y_train)
```

Out[23]:

```
GridSearchCV(cv=2, error_score='raise-deprecating', estimator=SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0, decision_function_shape='ovr', degree=3, gamma='auto_deprecated', kernel='rbf', max_iter=-1, probability=False, random_state=None, shrinking=True, tol=0.001, verbose=False), fit_params=None, iid='warn', n_jobs=None, param_grid={'gamma': [0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1]}, pre_dispatch='2*n_jobs', refit=True, return_train_score='warn',
```

```
In [24]:
#для SVC
clf_gs_svm_svc.best_params_
Out[24]:
{'gamma': 0.9}
In [25]:
parameters = {\footnote{\text{min_impurity_decrease}}: [0.9,0.8,0.7,0.6,0.5,0.4,0.3,0.2,0.1]}
clf_gs_decision_tree = GridSearchCV(DecisionTreeClassifier(), parameters, cv=2, scoring='accuracy')
clf_gs_decision_tree.fit(X_train, y_train)
Out[25]:
GridSearchCV(cv=2, error_score='raise-deprecating',
    estimator=DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=None,
       max_features=None, max_leaf_nodes=None,
       min_impurity_decrease=0.0, min_impurity_split=None,
       min_samples_leaf=1, min_samples_split=2,
       min_weight_fraction_leaf=0.0, presort=False, random_state=None,
       splitter='best'),
    fit_params=None, iid='warn', n_jobs=None,
    param_grid={'min_impurity_decrease': [0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1]},
    pre_dispatch='2*n_jobs', refit=True, return_train_score='warn',
    scoring='accuracy', verbose=0)
In [26]:
#для дерева решений
clf_gs_decision_tree.best_params_
Out[26]:
{'min_impurity_decrease': 0.4}
Для найденных оптимальных значений:
In [27]:
sgd2 = SGDClassifier(alpha=0.1).fit(X_train, y_train)
C:\Users\Dovlat\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:166: FutureWarning: max_iter and tol parameters have be
en added in SGDClassifier in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=10
00. From 0.21, default max_iter will be 1000, and default tol will be 1e-3.
 FutureWarning)
In [28]:
target_sgd2= sgd2.predict(X_test)
In [29]:
accuracy_score(y_test, target_sgd2), \
precision_score(y_test, target_sgd2), \
recall_score(y_test, target_sgd2)
Out[29]:
(0.7049180327868853, 0.6666666666666666, 0.8387096774193549)
In [30]:
svc2 = SVC(gamma=0.9).fit(X_train, y_train)
In [31]:
```

scoring - accuracy, versuse-u

target_svc2 = svc2.predict(X_test)

	_
In [32]:	
<pre>accuracy_score(y_test, target_svc2), \ precision_score(y_test, target_svc2), \ recall_score(y_test, target_svc2)</pre>	•
Out[32]:	
(0.5081967213114754, 0.5081967213114754, 1.0)	
In [33]:	
tree2 = DecisionTreeClassifier(random_state=1, min_impurity_decrease=0.4, max_depth=0.75).fit(X_train, y_train)	
In [34]:	
target_tree2 = tree2.predict(X_test)	
In [35]:	
<pre>accuracy_score(y_test, target_tree2), \ precision_score(y_test, target_tree2), \ recall_score(y_test, target_tree2)</pre>	•
Out[35]:	
(0.5081967213114754, 0.5081967213114754, 1.0)	