



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ Информатика и системы управления

КАФЕДРА Системы обработки информации и управления

Отчёт по лабораторной работе №4

По дисциплине:
«Технологии машинного обучения»

Выполнил:

Студент группы ИУ5-61Б

(Подпись, дата)

Пахомкин К.С.

(Фамилия И.О.)

Проверил:

(Подпись, дата)

Гапанюк Ю. Е.

(Фамилия И.О.)

Москва, 2021

Задание:

1. Выберите набор данных (датасет) для решения задачи классификации или регрессии.
2. В случае необходимости проведите удаление или заполнение пропусков и кодирование категориальных признаков.
3. С использованием метода `train_test_split` разделите выборку на обучающую и тестовую.
4. Обучите следующие модели:
 - одну из линейных моделей;
 - SVM;
 - дерево решений.
5. Оцените качество моделей с помощью двух подходящих для задачи метрик. Сравните качество полученных моделей.

ЛР4

```
In [1]: import numpy as np
import pandas as pd
from typing import Dict, Tuple
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
from sklearn.impute import SimpleImputer
import warnings
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import PolynomialFeatures
from sklearn.metrics import confusion_matrix, precision_score, recall_score
from sklearn.linear_model import LinearRegression
warnings.simplefilter("ignore")
```

```
In [2]: # чтение обучающей выборки
data = pd.read_csv('insurance.csv')
data.head()
```

```
Out[2]:
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

```
In [3]: from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
```

```
In [4]: # Кодирование значений
le = LabelEncoder()
# "sex" - пол
le.fit(data.sex.drop_duplicates())
data.sex = le.transform(data.sex)
# "smoker" - курение
le.fit(data.smoker.drop_duplicates())
data.smoker = le.transform(data.smoker)
# "region" - регион
le.fit(data.region.drop_duplicates())
data.region = le.transform(data.region)
```

```
In [5]: # Построим корреляционную матрицу
fig, ax = plt.subplots(figsize=(15,7))
sns.heatmap(data.corr(method='pearson'), ax=ax, annot=True, fmt='.2f')
```

Out[5]: <AxesSubplot:>



```
In [6]: X = data.drop(['charges'], axis = 1)
Y = data.charges.astype(int)
print('Входные данные:\n\n', X.head(), '\n\nВыходные данные:\n\n', Y.head())
```

Входные данные:

	age	sex	bmi	children	smoker	region
0	19	0	27.900	0	1	3
1	18	1	33.770	1	0	2
2	28	1	33.000	3	0	2
3	33	1	22.705	0	0	1
4	32	1	28.880	0	0	1

Выходные данные:

0	16884
1	1725
2	4449
3	21984
4	3866

Name: charges, dtype: int64

```
In [7]: X_train, X_test, Y_train, Y_test = train_test_split(X, Y, random_state=0)
print('Входные параметры обучающей выборки:\n\n', X_train.head(), \
      '\n\nВходные параметры тестовой выборки:\n\n', X_test.head(), \
      '\n\nВыходные параметры обучающей выборки:\n\n', Y_train.head(), \
      '\n\nВыходные параметры тестовой выборки:\n\n', Y_test.head())
```

Входные параметры обучающей выборки:

	age	sex	bmi	children	smoker	region
461	42	1	30.00	0	1	3
322	34	1	30.80	0	1	3
224	42	1	24.64	0	1	2
711	50	0	23.54	2	0	2
58	53	0	22.88	1	1	2

Входные параметры тестовой выборки:

	age	sex	bmi	children	smoker	region
578	52	1	30.200	1	0	3
610	47	0	29.370	1	0	2
569	48	1	40.565	2	1	1
1034	61	1	38.380	0	0	1
198	51	0	18.050	0	0	1

Выходные параметры обучающей выборки:

```
461    22144
322    35491
224    19515
711    10107
58     23244
Name: charges, dtype: int64
```

Выходные параметры тестовой выборки:

```
578     9724
610     8547
569    45702
1034    12950
198     9644
Name: charges, dtype: int64
```

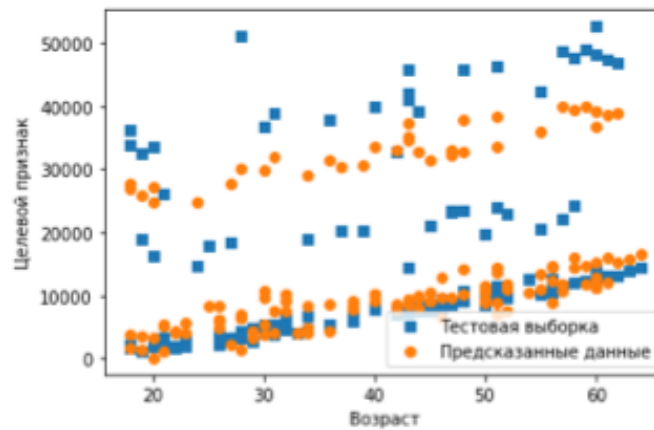
Построение линейной регрессии

```
In [8]: from sklearn.linear_model import LinearRegression
        from sklearn.metrics import mean_absolute_error, mean_squared_error, median_absolute_error
```

```
In [9]: Lin_Reg = LinearRegression().fit(X_train, Y_train)

        lr_y_pred = Lin_Reg.predict(X_test)
```

```
In [10]: plt.scatter(X_test.age, Y_test, marker = 's', label = 'Тестовая выборка')
         plt.scatter(X_test.age, lr_y_pred, marker = 'o', label = 'Предсказанные данные')
         plt.legend (loc = 'lower right')
         plt.xlabel ('Возраст')
         plt.ylabel ('Целевой признак')
         plt.show()
```

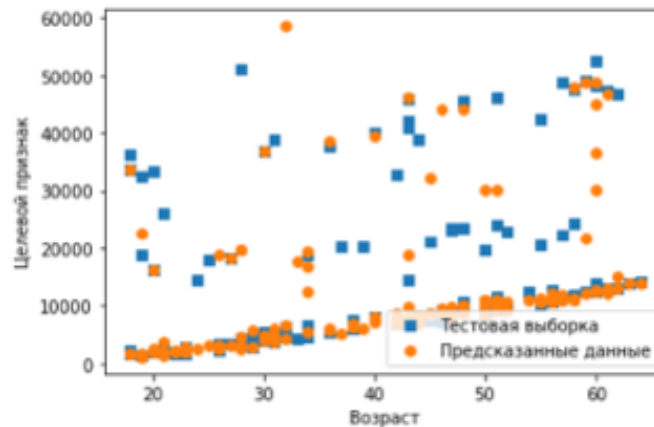


SVM

```
In [11]: from sklearn.svm import SVC, LinearSVC
from matplotlib import pyplot as plt
```

```
In [12]: svc = SVC(kernel='linear')
svc.fit(X_train, Y_train)
lr_y_pred = svc.predict(X_test)
```

```
In [13]: plt.scatter(X_test.age, Y_test, marker='s', label='Тестовая выборка')
plt.scatter(X_test.age, lr_y_pred, marker='o', label='Предсказанные данные')
plt.legend(loc='lower right')
plt.xlabel('Возраст')
plt.ylabel('Целевой признак')
plt.show()
```



Tree

```
In [14]: from sklearn.tree import DecisionTreeClassifier, DecisionTreeRegressor, export_graphviz
from sklearn.tree import export_graphviz
from sklearn import tree
import re
```

```
In [15]: # Обучим дерево на всех признаках iris
clf = tree.DecisionTreeClassifier()
clf = clf.fit(X,Y)
lr_y_pred = clf.predict(X_test)
```

```
In [16]: from IPython.core.display import HTML
from sklearn.tree.export import export_text
tree_rules = export_text(clf, feature_names=list(X.columns))
HTML('<pre>' + tree_rules + '</pre>')
```

```
Out[16]: |--- age <= 18.50
|         |--- bmi <= 39.15
|         |         |--- bmi <= 37.73
|         |         |         |--- bmi <= 33.99
|         |         |         |         |--- bmi <= 23.15
|         |         |         |         |         |--- bmi <= 22.39
|         |         |         |         |         |         |--- children <= 1.00
|         |         |         |         |         |         |         |--- sex <= 0.50
|         |         |         |         |         |         |         |         |--- region <= 1.00
|         |         |         |         |         |         |         |         |         |--- class: 14283
|         |         |         |         |         |         |         |         |         |--- region > 1.00
|         |         |         |         |         |         |         |         |         |         |--- class: 1607
|         |         |         |         |         |         |         |         |         |--- sex > 0.50
|         |         |         |         |         |         |         |         |         |         |--- smoker <= 0.50
|         |         |         |         |         |         |         |         |         |         |         |--- bmi <= 18.71
|         |         |         |         |         |         |         |         |         |         |         |         |--- class: 1694
|         |         |         |         |         |         |         |         |         |         |         |         |--- bmi > 18.71
|         |         |         |         |         |         |         |         |         |         |         |         |         |--- class: 1702
|         |         |         |         |         |         |         |         |         |         |         |         |         |--- smoker > 0.50
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- class: 13747
|         |         |         |         |         |         |         |         |         |         |         |         |         |--- children > 1.00
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- smoker <= 0.50
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- class: 11884
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- smoker > 0.50
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- class: 12829
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- bmi > 22.39
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- class: 1704
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- bmi > 23.15
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- bmi <= 26.15
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- bmi <= 25.32
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- children <= 0.50
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- sex <= 0.50
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- class: 2196
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- sex > 0.50
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- bmi <= 23.48
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- class: 1121
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- bmi > 23.48
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- bmi <= 24.46
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- class: 1705
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- bmi > 24.46
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- class: 15518
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- children > 0.50
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- sex <= 0.50
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |--- class: 2201
```

