

Лабораторная работа №5
по дисциплине
«Технологии машинного обучения»
на тему
«Линейные модели, SVM и деревья решений»

Выполнил:
студент группы ИУ5-61Б
Белоусов Е. А.

1. Лабораторная работа №5

1.1. Линейные модели, SVM и деревья решений.

1.2. Цель лабораторной работы:

изучение линейных моделей, SVM и деревьев решений.

1.3. 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.

```
[1]: import numpy as np
import pandas as pd
import sklearn
```

```
[2]: data = pd.read_csv('../data/winequality-red.csv')
data.head()
```

```
[2]: fixed acidity volatile acidity citric acid residual sugar chlorides \
0      7.4          0.70      0.00      1.9    0.076
1      7.8          0.88      0.00      2.6    0.098
2      7.8          0.76      0.04      2.3    0.092
3     11.2          0.28      0.56      1.9    0.075
4      7.4          0.70      0.00      1.9    0.076
```

```
free sulfur dioxide total sulfur dioxide density pH sulphates \
0         11.0          34.0 0.9978 3.51    0.56
1         25.0          67.0 0.9968 3.20    0.68
2         15.0          54.0 0.9970 3.26    0.65
3         17.0          60.0 0.9980 3.16    0.58
4         11.0          34.0 0.9978 3.51    0.56
```

```
alcohol quality
0    9.4      5
1    9.8      5
2    9.8      5
3    9.8      6
4    9.4      5
```

```
[3]: data.shape
```

```
[3]: (1599, 12)
```

```
[5]: data.isna().sum()
```

```
[5]: fixed acidity      0
      volatile acidity  0
      citric acid       0
      residual sugar    0
      chlorides         0
      free sulfur dioxide 0
      total sulfur dioxide 0
      density          0
      pH              0
      sulphates        0
      alcohol          0
      quality          0
      dtype: int64
```

```
[6]: y = data['quality']
```

```
[7]: del data['quality']
```

```
[9]: from sklearn.model_selection import train_test_split
      X_train, X_test, y_train, y_test = train_test_split(data.values, y, test_size=0.2, random_state=42)
      X_train.shape, X_test.shape
```

```
[9]: ((1279, 11), (320, 11))
```

```
[10]: from sklearn.metrics import mean_squared_error, mean_absolute_error
      def score(result, y_test):
          print('MSE = ', mean_squared_error(result, y_test))
          print('MAE = ', mean_absolute_error(result, y_test))
```

```
[14]: from sklearn.linear_model import LinearRegression
      lr = LinearRegression()
      lr.fit(X_train, y_train)
      result = lr.predict(X_test)
      score(result, y_test)
```

```
MSE = 0.39002514396395493
MAE = 0.5035304415524369
```

```
[16]: from sklearn.svm import SVR
      svr = SVR()
      svr.fit(X_train, y_train)
      score(svr.predict(X_test), y_test)
```

```
MSE = 0.5325010802600723
MAE = 0.5657291310543899
```

```
[17]: from sklearn.tree import DecisionTreeRegressor
      tree = DecisionTreeRegressor()
      tree.fit(X_train, y_train)
```

```
score(tree.predict(X_test), y_test)
```

MSE = 0.6125

MAE = 0.4625

С настройками по умолчанию, лучший результат по оценке MSE у линейной модели. Лучший результат по оценке MAE у модели решающего дерева.