Лабораторная работа №6. Градиентные методы в решении задач машинного обучения.

Часть 1. Линейная регрессия.

Используемый набор данных: [Concrete Compressive Strength (https://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength)](https://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength).

In [1]:

```
from IPython.display import display
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
import os
import requests
import xlrd # для pd.read_excel()

%matplotlib inline

pd.options.display.max_columns = None
```

In [2]:

```
def downloadFile(url, filePath):
    if not os.path.exists(filePath):
        req = requests.get(url)
        f = open(filePath, "wb")
        f.write(req.content)
        f.close

url = "https://archive.ics.uci.edu/ml/machine-learning-databases/concrete/compressive"
downloadFile(url + "/Concrete_Data.xls", "dataset/Concrete_Data.xls")
downloadFile(url + "/Concrete_Readme.txt", "dataset/Concrete_Readme.txt")
```

In [3]:

```
headers = ["Cement", "Blast Furnace Slag", "Fly Ash", "Water", "Superplasticizer", "Coa
rse Aggregate", "Fine Aggregate",
           "Age", "Concrete compressive strength"]
data = pd.read_excel("dataset/Concrete_Data.xls", names=headers)
data.sample(40)
```

Out[3]:

	Cement	Blast Furnace Slag	Fly Ash	Water	Superplasticizer	Coarse Aggregate	Fine Aggregate	Age	Con compre stre
36	237.50	237.50	0.00	228.00	0.00	932.0	594.00	28	30.07
237	213.76	98.06	24.52	181.74	6.65	1066.0	785.52	56	47.13
137	362.60	189.00	0.00	164.90	11.60	944.7	755.80	28	71.29
617	254.00	0.00	0.00	198.00	0.00	968.0	863.00	3	9.30
101	388.60	97.10	0.00	157.90	12.10	852.1	925.70	7	34.90
25	380.00	0.00	0.00	228.00	0.00	932.0	670.00	270	53.30
768	331.00	0.00	0.00	192.00	0.00	978.0	825.00	180	38.99
334	275.07	0.00	121.35	159.48	9.90	1053.6	777.50	3	23.80
443	194.68	0.00	100.52	170.17	7.48	998.0	901.80	28	37.26
506	491.00	26.00	123.00	201.00	3.93	822.0	699.00	56	61.85
439	173.81	93.37	159.90	172.34	9.73	1007.2	746.60	28	37.81
988	153.60	144.20	112.30	220.10	10.10	923.2	657.90	28	16.50
494	387.00	20.00	94.00	157.00	14.32	938.0	845.00	56	56.33
206	212.07	0.00	121.62	180.31	5.69	1057.6	779.32	28	24.90
383	451.00	0.00	0.00	165.00	11.25	1030.0	745.00	28	78.80
462	172.38	13.61	172.37	156.76	4.14	1006.3	856.40	100	37.67
698	203.50	305.30	0.00	203.50	0.00	963.4	630.00	28	41.68
168	469.00	117.20	0.00	137.80	32.20	852.1	840.50	91	70.69
645	203.50	305.30	0.00	203.50	0.00	963.4	630.00	7	19.53
107	323.70	282.80	0.00	183.80	10.30	942.7	659.90	7	49.80
351	213.50	0.00	174.24	154.61	11.66	1052.3	775.48	28	45.93
922	255.00	99.00	77.00	189.00	6.00	919.0	749.00	28	33.79
794	302.00	0.00	0.00	203.00	0.00	974.0	817.00	180	26.74
834	310.00	143.00	111.00	168.00	22.00	914.0	651.00	28	33.68
409	167.35	129.90	128.62	175.46	7.79	1006.3	746.60	3	14.94
124	388.60	97.10	0.00	157.90	12.10	852.1	925.70	28	50.69
934	184.00	86.00	190.00	213.00	6.00	923.0	623.00	28	22.93
543	255.00	0.00	0.00	192.00	0.00	889.8	945.00	7	10.22
7	380.00	95.00	0.00	228.00	0.00	932.0	594.00	28	36.44
83	362.60	189.00	0.00	164.90	11.60	944.7	755.80	3	35.30
164	425.00	106.30	0.00	153.50	16.50	852.1	887.10	91	65.19
50	332.50	142.50	0.00	228.00	0.00	932.0	594.00	180	39.77
726	331.00	0.00	0.00	192.00	0.00	1025.0	821.00	3	14.30
27	342.00	38.00	0.00	228.00	0.00	932.0	670.00	180	52.12
135	439.00	177.00	0.00	186.00	11.10	884.9	707.90	28	65.99
468	213.50	0.00	174.24	159.21	11.66	1043.6	771.90	100	52.95

	Cement	Blast Furnace Slag	Fly Ash	Water	Superplasticizer	Coarse Aggregate	Fine Aggregate	Age	Con compre stre
712	192.00	288.00	0.00	192.00	0.00	929.8	716.10	7	21.48
661	141.30	212.00	0.00	203.50	0.00	971.8	748.50	7	10.39
987	162.00	190.10	148.10	178.80	18.80	838.1	741.40	28	33.76
739	296.00	0.00	0.00	186.00	0.00	1090.0	769.00	28	25.17

In [4]:

```
display(data.describe())
display(data.isna().sum())
```

```
Cement                         0
Blast Furnace Slag             0
Fly Ash                        0
Water                          0
Superplasticizer               0
Coarse Aggregate               0
Fine Aggregate                 0
Age                            0
Concrete compressive strength  0
dtype: int64
```

	Cement	Blast Furnace Slag	Fly Ash	Water	Superplasticizer	Coarse Aggregate	
count	1030.000000	1030.000000	1030.000000	1030.000000	1030.000000	1030.000000	10
mean	281.165631	73.895485	54.187136	181.566359	6.203112	972.918592	7
std	104.507142	86.279104	63.996469	21.355567	5.973492	77.753818	
min	102.000000	0.000000	0.000000	121.750000	0.000000	801.000000	5
25%	192.375000	0.000000	0.000000	164.900000	0.000000	932.000000	7
50%	272.900000	22.000000	0.000000	185.000000	6.350000	968.000000	7
75%	350.000000	142.950000	118.270000	192.000000	10.160000	1029.400000	8
max	540.000000	359.400000	200.100000	247.000000	32.200000	1145.000000	9

Пропусков в данных нет.

Подготовим выборки и обучим модель.

In [5]:

```
y = data["Concrete compressive strength"].copy()
X = data.drop(columns=["Concrete compressive strength"]).copy()

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.35, random_state=
25)
y_pred = LinearRegression().fit(X_train, y_train).predict(X_test)
```

Так как нельзя построить ROC-кривую для линейной регрессии (https://stats.stackexchange.com/questions/88920/how-to-create-roc-curve-to-assess-the-performance-of-regression-models) (в данном случае целевая переменная набора данных *"Прочность бетона на сжатие"* не предрасположена бинарной классификации, для которой предназначена ROC-кривая), оценим модель с помощью метода RMSE и коэффициента детерминации.

In [6]:

```
RMSE = mean_squared_error(y_test, y_pred, squared=False)
r2 = r2_score(y_test, y_pred)
print("RMSE = {0:0.3f}\nr2 = {1:0.3f}".format(RMSE, r2))
```

```
RMSE = 10.102
r2 = 0.622
```

Диаграмма рассеяния для построенной модели.

In [7]:

```
_, ax = plt.subplots()
ax.scatter(y_test, y_pred, s = 5, color = "r", alpha = 0.75)
ax.set_title("Scatter Plot")
ax.set_xlabel("Concrete compressive strength")
ax.set_ylabel("Predicted")
display()
```