Шмаков Данила 181-351

# Задание 7

## Для этого задания используется значительно большее количество признаков

```
data = pd.read_csv('IoT.csv', delimiter=',')
quantity = ['ts', 'duration', 'orig_bytes', 'orig_pkts', 'resp_bytes', 'resp_pkts', 'resp_ip_bytes', 'missed_bytes']
quality = ['proto', 'conn_state', 'service', 'Local_orig', 'Local_resp', 'History']
data = pd.DataFrame(data, columns=quantity+quality)
```

ПредПредобработка данных такая же как и в заданиях 3-5 (заменяем '-' на среднее значение). Также в отличие от 3-5 заменяем nan на 0.

Кодируем качественные в количественные признаки

```
le = LabelEncoder()

for item in quality:
    data[item]= le.fit_transform(data[item])
    dtype = float
```

Нормируем

```
ss = StandardScaler()
data.iloc[:, :-1] = ss.fit_transform(data.iloc[:,:-1])
```

Создаем объект метода главных компонент

```
pca = PCA(svd_solver='full')
```

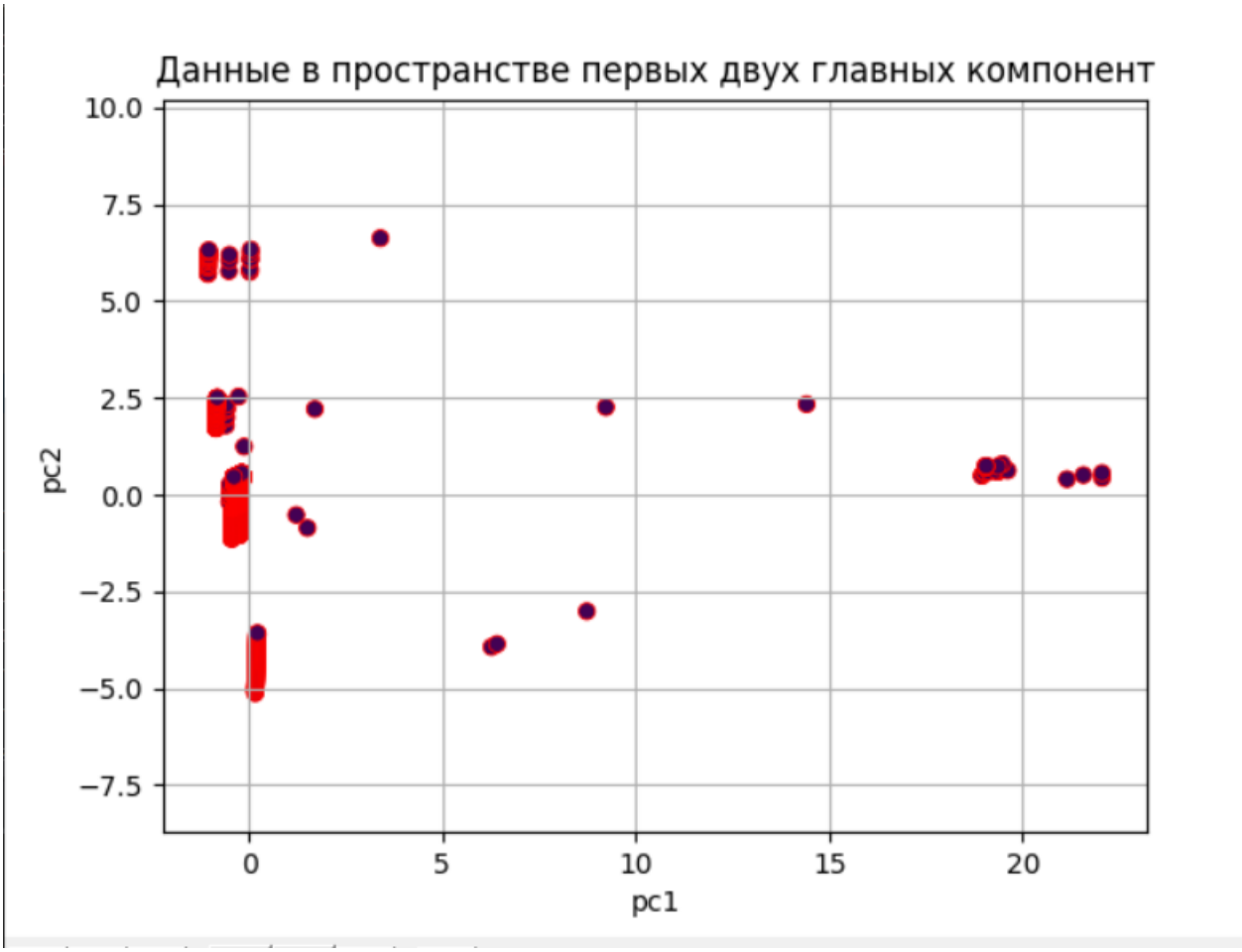
Используем метод главных компонент

```
datadf = data.copy()

column=[]
for i in range(len(data.columns) -1):
    column.append(f'pc{i+1}')
column.append('label')
datadf.columns = column
datadf.iloc[:,:-1] = pca.fit_transform(datadf.iloc[:, :-1])
```

Рисуем данные в пространстве главных компонент

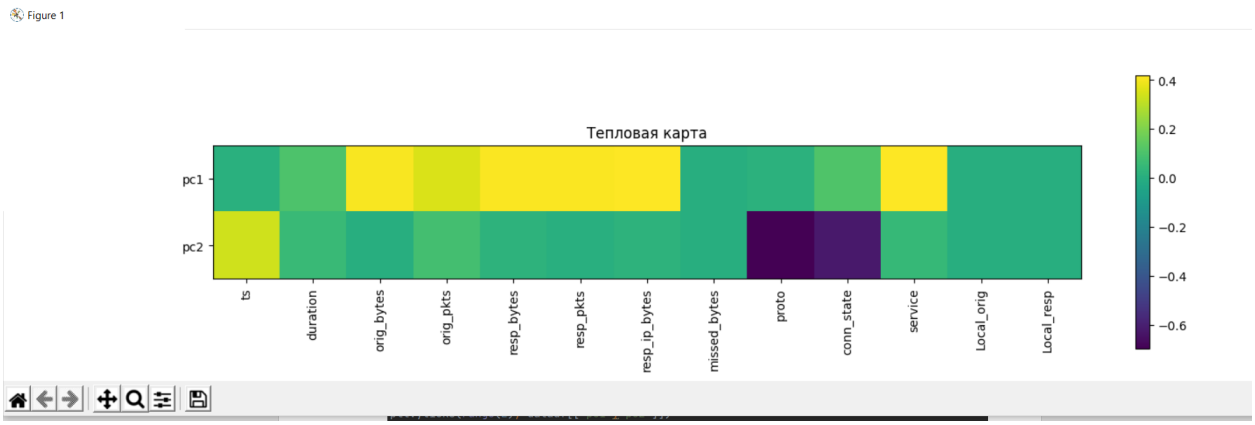
```
#Рисуем данные в пространстве первых двух главных компонент
plt.figure()
plt.grid()
plt.scatter(datadf['pc1'], datadf['pc2'], c=le.fit_transform(datadf['label']), lw=.6, edgecolors='red')
plt.axis('equal')
plt.title("Данные в пространстве первых двух главных компонент")
plt.xlabel("pc1")
plt.ylabel("pc2")
plt.show()
```




Рисуем тепловую карту

```
#Рисуем тепловую карту по координатам первых двух главных компонент
plt.matshow(pca.components_[:2])
plt.colorbar()
plt.gca().xaxis.tick_bottom()
plt.xticks(range(len(datadf.columns) -1), data.iloc[:,:-1], rotation=90)
plt.yticks(range(2), datadf[['pc1','pc2']])
plt.title("Тепловая карта")

plt.show()
```




Можно заметить, что признаки missed_bytes, Local_orig, Local_resp наименее информативны, так как значения близки к нулю. Признаки orig_bytes/pkts и resp_bytes/pkts, pesp_ip_bytes, service – наиболее информативны так как их значение наиболее отлично от нуле по первой компоненте.