

Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧЕРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
“НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО”
Факультет ПИиКТ



ОТЧЁТ
По лабораторной работе №5
По предмету: Системы искусственного интеллекта
Вариант: 1

Студент:
Андрейченко Леонид Вадимович
Группа Р33301

Преподаватель:
Кугаевских Александр Владимирович

Санкт – Петербург

2022

Количество главных компонент

Минимально количество главных компонент, которое необходимо для того, чтобы доля объясненной дисперсии была больше чем 0.81 = 46

```
[37] from sklearn.decomposition import PCA

pca = PCA(n_components=46, svd_solver='full')
modelPCA = pca.fit(X_train)

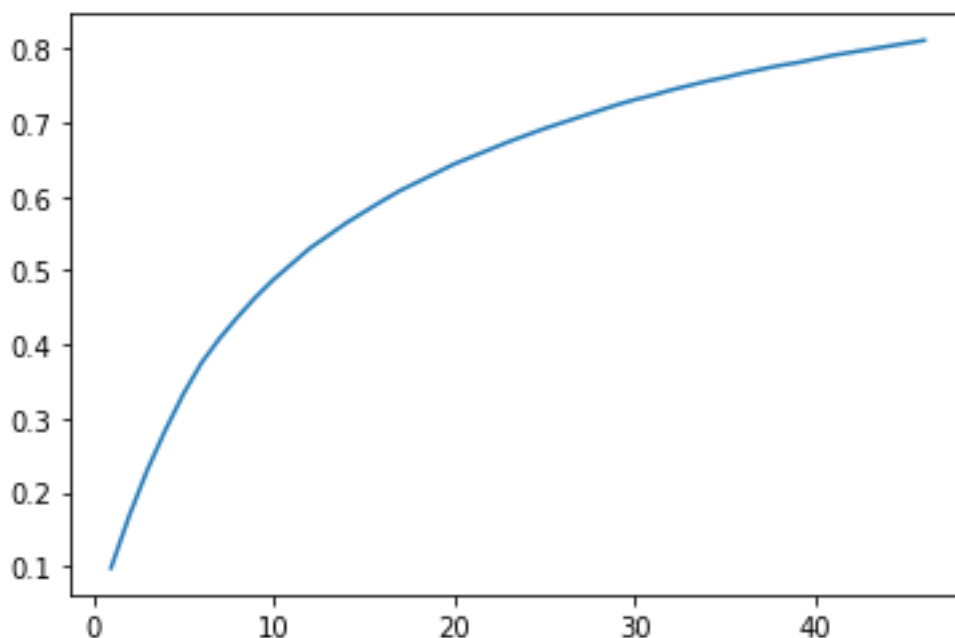
X_train = modelPCA.transform(X_train)
```

Найдем долю объяснённой дисперсии в зависимости от количества используемых ГК.

```
explained_variance = np.round(np.cumsum(pca.explained_variance_ratio_),3)
explained_variance
```

```
array([0.098, 0.168, 0.23 , 0.284, 0.333, 0.375, 0.408, 0.437, 0.464,
       0.488, 0.509, 0.53 , 0.547, 0.564, 0.579, 0.594, 0.608, 0.62 ,
       0.632, 0.644, 0.654, 0.664, 0.674, 0.683, 0.692, 0.7 , 0.708,
       0.716, 0.724, 0.731, 0.737, 0.744, 0.75 , 0.756, 0.761, 0.767,
       0.772, 0.777, 0.781, 0.786, 0.791, 0.795, 0.799, 0.803, 0.807,
       0.811])
```

График зависимости доли объясненной дисперсии от количества ГК



Количество верно классифицированных объектов класса 0

```
CM[0][0]
```

1376

Вероятность отнесения 5 объектов к верному классу

```
[60] sum_r = 0
     diag = 0
     for i in range(10):
         diag += CM[i][i]
         sum_r += sum(CM[i])
     ver = (diag/sum_r)**5
     ver
```

0.08714698014712174

Метрики модели

	precision	recall	f1-score	support
0	0.796	0.813	0.804	1693
1	0.911	0.885	0.898	2075
2	0.386	0.492	0.432	1763
3	0.662	0.797	0.723	1873
4	0.666	0.755	0.708	1756
5	0.466	0.439	0.452	1591
6	0.449	0.387	0.416	1766
7	0.740	0.680	0.708	1886
8	0.370	0.332	0.350	1773
9	0.635	0.492	0.554	1824
accuracy			0.614	18000
macro avg	0.608	0.607	0.605	18000
weighted avg	0.615	0.614	0.611	18000

Выводы

По полученным метрикам видно, что в целом модель получилась не очень точной. Из этого заключения можно сделать вывод о том, что метод главных компонент не всегда является достаточно эффективным для выбора значимых факторов данных.