Брагин Алексей. КЭ - 402

In [ ]:
```python
import numpy as np
import pandas as pd
from sklearn import datasets
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from mpl_toolkits.mplot3d import Axes3D
```

In [2]:
```python
np.random.seed(5)

iris = datasets.load_iris()
X = iris.data
y = iris.target
```

Кластеризация методом k-means с количеством кластеров 3

Как видно данные хорошо разделяются на 3 кластера и совпадают с "настоящими" значениями

In [3]:
```python
fig = plt.figure(1, figsize=(8, 6))
ax = Axes3D(fig, rect=[0, 0, .95, 1], elev=48, azim=134)

kmeans = KMeans(n_clusters=3).fit(X)
labels = kmeans.labels_

ax.scatter(X[:, 3], X[:, 0], X[:, 2],
           c=labels.astype(float), edgecolor='k')

ax.w_xaxis.set_ticklabels([])
ax.w_yaxis.set_ticklabels([])
ax.w_zaxis.set_ticklabels([])
ax.set_xlabel('Petal width')
ax.set_ylabel('Sepal length')
ax.set_zlabel('Petal length')
ax.set_title('kmeans - 3 clusters')
ax.dist = 12

fig.show()

# Plot the ground truth
fig = plt.figure(2, figsize=(8, 6))
ax = Axes3D(fig, rect=[0, 0, .95, 1], elev=48, azim=134)

for name, label in [('Setosa', 0),
                    ('Versicolour', 1),
                    ('Virginica', 2)]:
    ax.text3D(X[y == label, 3].mean(),
              X[y == label, 0].mean(),
              X[y == label, 2].mean() + 2, name,
              horizontalalignment='center',
              bbox=dict(alpha=.2, edgecolor='w', facecolor='w'))
# Reorder the labels to have colors matching the cluster results
y = np.choose(y, [1, 2, 0]).astype(float)
ax.scatter(X[:, 3], X[:, 0], X[:, 2], c=y, edgecolor='k')
```

```python
ax.w_xaxis.set_ticklabels([])
ax.w_yaxis.set_ticklabels([])
ax.w_zaxis.set_ticklabels([])
ax.set_xlabel('Petal width')
ax.set_ylabel('Sepal length')
ax.set_zlabel('Petal length')
ax.set_title('Ground Truth')
ax.dist = 12

fig.show()
```
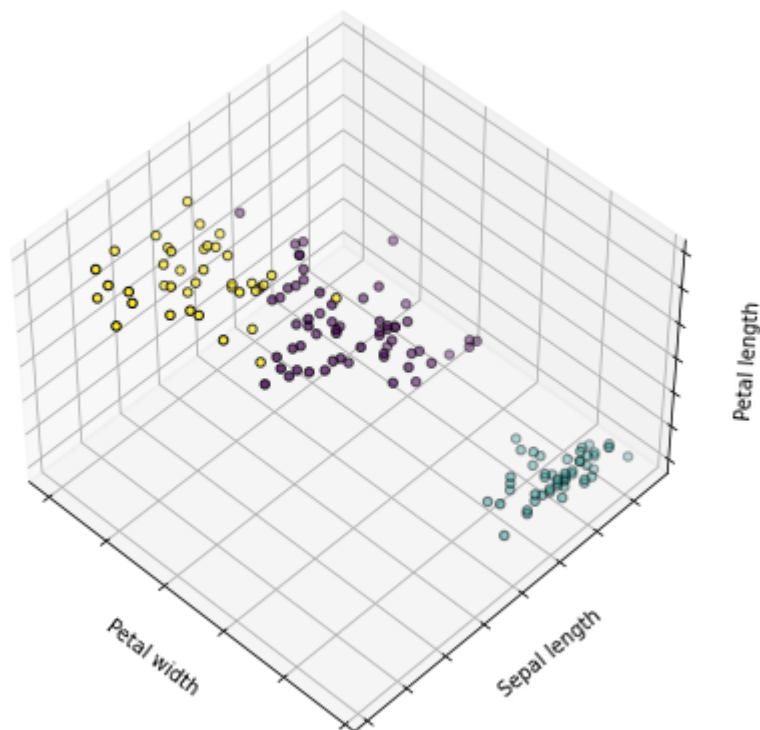
```
<ipython-input-3-58189091760b>:19: UserWarning: Matplotlib is currently using mo
dule://ipykernel.pylab.backend_inline, which is a non-GUI backend, so cannot sho
w the figure.
  fig.show()
<ipython-input-3-58189091760b>:46: UserWarning: Matplotlib is currently using mo
dule://ipykernel.pylab.backend_inline, which is a non-GUI backend, so cannot sho
w the figure.
  fig.show()
```
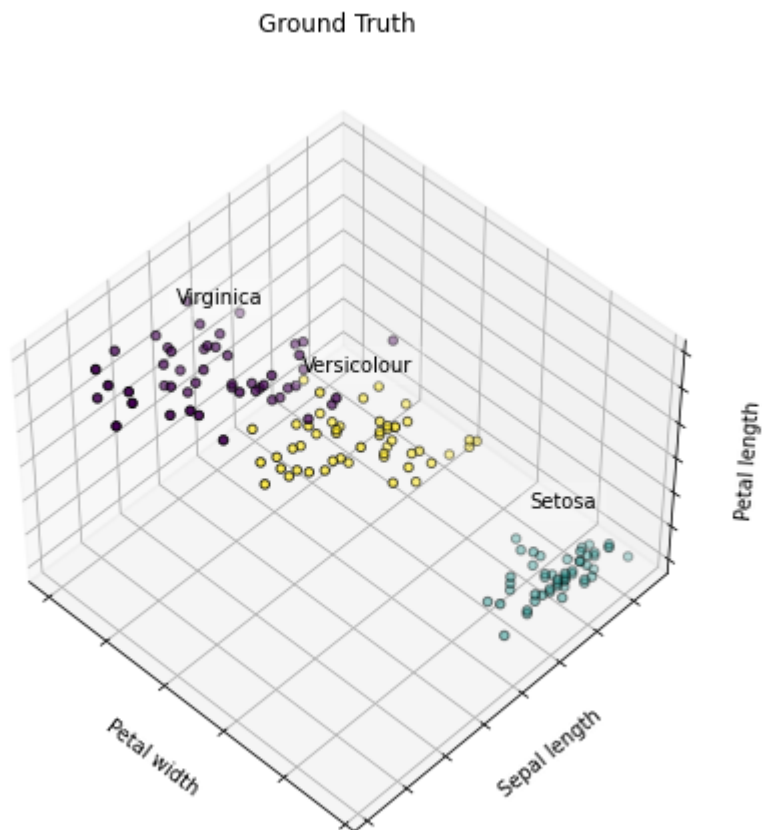

kmeans - 3 clusters

Ground Truth


```
In [14]:  customers_data = pd.read_csv('customers.csv')
          customers_data = customers_data.drop(['Row', 'CustomerId'], axis=1)
```

```
In [15]:  customers_data.head()
```

Out[15]:

	Age	Education	YearsEmployed	Income	CardDebt	OtherDebt	Defaulted	DebtIncomeRatio
0	41	2	6	19	0.124	1.073	0.0	6.3
1	47	1	26	100	4.582	8.218	0.0	12.8
2	33	2	10	57	6.111	5.802	1.0	20.9
3	29	2	4	19	0.681	0.516	0.0	6.3
4	47	1	31	253	9.308	8.908	0.0	7.2

```
In [24]:  customers_data = customers_data.dropna()
```

Кластеризация данных о покупателях методом k-means с количеством кластеров 8

На данных отчетливо видно, что на протяжениий всех возрастов доход зависит от полученного образования

```
In [39]:  kmeans = KMeans(n_clusters=8).fit(customers_data)
```

```python
labels = kmeans.labels_

fig = plt.figure(1, figsize=(8, 6))
ax = Axes3D(fig, rect=[0, 0, .95, 1], elev=48, azim=134)

ax.scatter(customers_data['Age'], customers_data['Education'], customers_data['I
          c=labels.astype(float), edgecolor='k')
ax.w_xaxis.set_ticklabels([])
ax.w_yaxis.set_ticklabels([])
ax.w_zaxis.set_ticklabels([])
ax.set_xlabel('Age')
ax.set_ylabel('Education')
ax.set_zlabel('Income')
ax.set_title('kmeans - 8 clusters')
ax.dist = 12

fig.show()
```
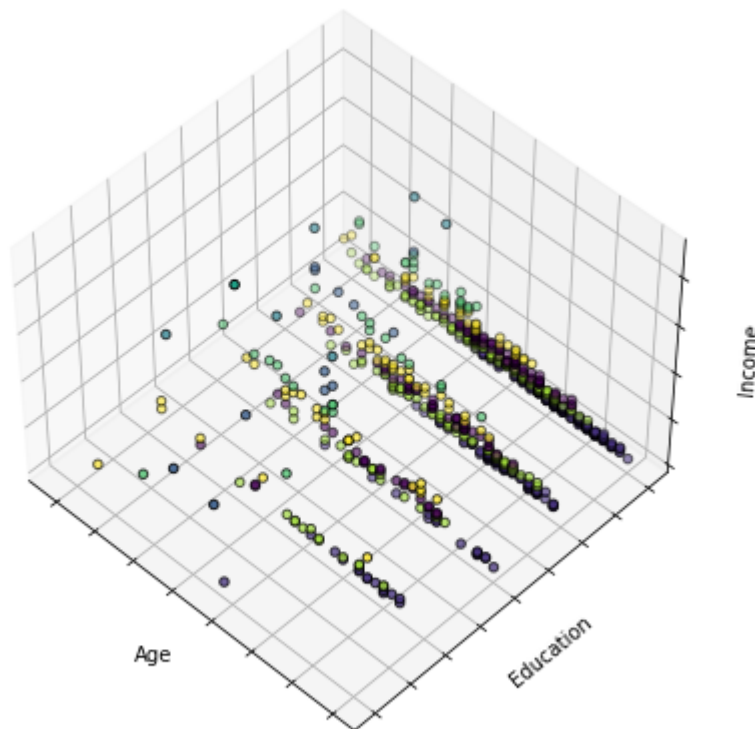
```
<ipython-input-39-1cb068130d95>:18: UserWarning: Matplotlib is currently using m
odule://ipykernel.pylab.backend_inline, which is a non-GUI backend, so cannot sh
ow the figure.
  fig.show()
```


kmeans - 8 clusters

In [ ]:

In [ ]:

In [ ]:

In [ ]: