# Понижение размерности с методом t-SNE

Метод t-SNE (t-distributed stochastic neighbor embedding) представляет собой один из методов обучения без учителя, используемых для визуализации, например, отображения пространства высокой размерности в двух- или трехмерное пространство. t-SNE расшифровывается как распределенное стохастическое соседнее вложение.

Метод моделирует каждый объект пространства высокой размерности в двух- или трехкоординатную точку таким образом, что близкие по характеристикам элементы данных в многомерном пространстве (например, датасете с большим числом столбцов) проецируются в соседние точки, а разнородные объекты с большей вероятностью моделируются точками, далеко отстоящими друг от друга.

Вернемся к примеру с ирисами и посмотрим, как произвести моделирование по этому методу при помощи библиотеки sklearn.

# Импорт библиотек

from sklearn import datasets

from sklearn.manifold import TSNE

import matplotlib.pyplot as plt

# Загрузка датасета

iris\_df = datasets.load\_iris()

# Определяем модель и скорость обучения

model = TSNE(learning\_rate=100)

# Обучаем модель

transformed = model.fit\_transform(iris\_df.data)

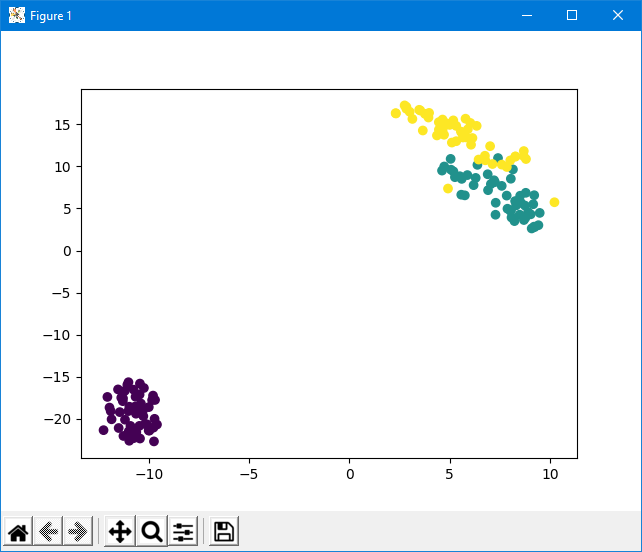
# Представляем результат в двумерных координатах

x\_axis = transformed[:, 0]

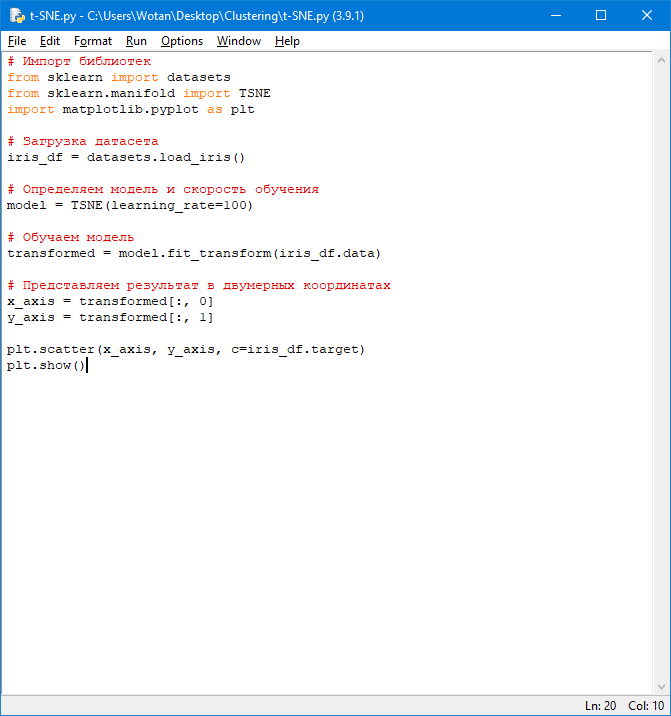
y\_axis = transformed[:, 1]

plt.scatter(x\_axis, y\_axis, c=iris\_df.target)

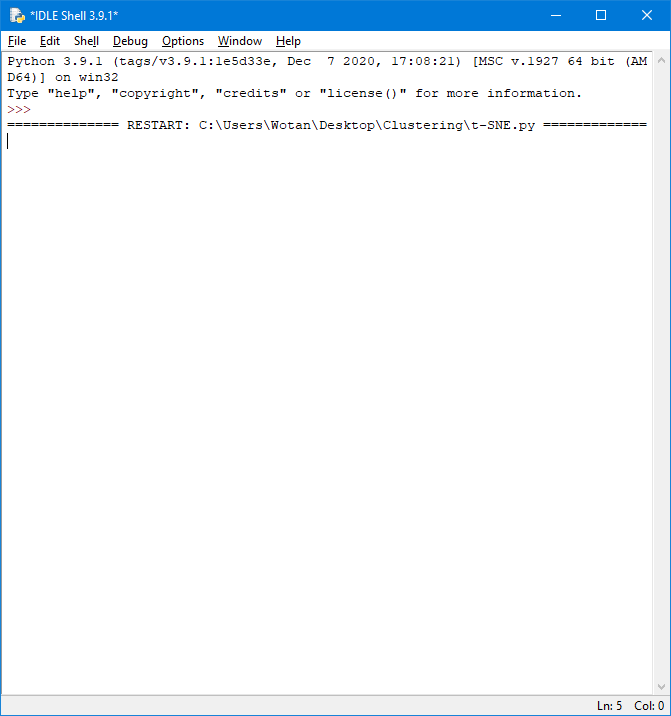
plt.show()



Запуск программы



t-SNE.py



Python терминал