## Классификация изображений

### Цель работы

Применить методы машинного обучения для решения задач классификации изображений.

### Задания для выполнения

1. Загрузите датасет Olivetti faces;
2. Познакомьтесь с описанием и структурой датасета. Описание можно найти в [документации](https://scikit-learn.org/stable/datasets/index.html#real-world-datasets).
3. Выведите информацию о количественных параметрах датасета;
4. Выведите несколько изображений на экран используя инструментарий библиотеки matplotlib;
5. Разделите эти данные на тестовую и обучающую выборки;
6. Постройте модель классификатора метода опорных векторов для идентификации человека по изображению;
7. Оцените качество модели на тестовой выборке с помощью следующих метрик:
   1. достоверность предсказания (accuracy);
   2. точность (precision);
   3. полнота (recall);
8. Постройте кривую обучения - график зависимости тестовой и обучающей эффективности от размера обучающей выборки.
9. Сделайте вывод о применимости модели.

### Методические указания

Датасет Olivetti faces - это один из известных модельных наборов данных для обучения методам классификации изображений. Его можно получить с помощью стандартных средств sklearn:

from sklearn.datasets import fetch\_olivetti\_faces

faces = fetch\_olivetti\_faces()

print(faces.DESCR)

При работе с разными данными часто требуется выводить из на экран. Хотя это абсолютно не нужно для работы самих алгоритмов машинного обучения, зачастую проводится, например, анализ ошибок. Для этого нужно уметь выводить изображения стандартными средствами. В библиотеке matplotlib, например, можно воспользоваться функцией imshow:

p.imshow(images[i], cmap=plt.cm.bone)

Более подробно про эту функцию можно прочитать в [документации](https://matplotlib.org/3.3.2/api/_as_gen/matplotlib.pyplot.imshow.html).

Методы машинного обучения библиотеки sklearn рассчитаны, что на вход будет подаваться плоский массив признаков. Однако, изображение - это двумерный объект. Можно разными способами превратить изображение в набор признаков. В нейронных моделях для этого используются свертки. Но самый простой способ - просто представить изображение как линейную последовательность пикселей.

Если изображение имеет разрешение 64 на 64 пикселя, то всего пикселей получается 4096. Если изображение черно-белое (в градациях серого), то значением признака, соответствующего каждому пикселю можно взять яркость этого пикселя. Если изображение цветное, то каждому пикселю может соответствовать три разных признака. Причем обычно такие значения нормируются к шкале долей единицы.

Такой набор признаков уже можно использовать как исходные данные для машинного обучения. Далее алгоритм машинного обучения полностью совпадает с работой с численными данными.

### Контрольные вопросы

1. Какие выводы мы можем сделать на основании метрик модели, построенной в данной лабораторной работе?
2. Как представляется изображение в методах машинного обучения? От чего зависит размерность вектора признаков?
3. Какие задачи можно решать при помощи анализа изображений? Какие из них самые распространенные?
4. Какие известные датасеты изображений существуют? Для каких задач они применяются?

### Дополнительные задания

1. Постройте модели классификации на основе следующих методов:
   1. логистическая регрессия (LogisticRegression);
   2. метод опорных векторов с гауссовым ядром (SVC);
   3. метод опорных векторов с полиномиальным ядром (SVC);
   4. метод k ближайших соседей (KNeighborsClassifier);
   5. многослойный перцептрон (MLP);
   6. другие методы по желанию;
2. Проанализируйте метрики каждой модели и сделайте выводы об их эффективности и применимости. Сравните эффективность всех этих моделей и выберите лучшую;
3. Для каждой модели из п.3 постройте кривые обучения и диагностируйте недо-/переобучение модели. Попробуйте изменить параметр регуляризации для улучшения результатов модели.
4. Сделайте замеры времени обучения для каждой модели. Сделайте вывод о сравнительной производительности моделей.
5. (\*) Используйте сверточную нейронную сеть для решения той же задачи. Сравните ее эффективность и производительность с классическими моделями.