Задание №2

1. Для систем контроля и статистики крайне актуальной является проблема точности определения показателей. Так например для системы пропуска сотрудников крайне актуальной является проблема определения является ли человек подошедший к проходной нашим сотрудником. В решении данной проблемы стремительно развиваются системы машинного обучения
2. Сегодня используется множество моделей машинного обучения способных решить данную задачу с большей или меньшей точностью. Работа этих моделей основана на расчете Эвклидовой дистанции, среднеквадратичного отклонения и др. методы позволяющие нам определить разницу меду двумя наборами данных, а именно данные в которых мы уверены, то есть наша база знаний и данные поступающие для проверки и сравнения.
3. Суть работы программы

А) загрузка базы с фотографиями и приведение их к оттенкам серого, это необходимо для уменьшения погрешности, из за влияния цветовой гаммы.

Б) поскольку у нас нет входных данных мы их получаем разделяя исходную выборку на 2 части - train и test.

В) используя функцию среднеквадратичного отклонения мы определяем нашу таблицу с данными о людях и командой pred = regr.predict(x\_test) рассчитываем совпадение двух наборов данных, r2\_score(y\_test, pred) дает нам % точности

4. На основании расчетов были получены следующие данные

При случайном распределении тестовых и тренировочных данных успешность расчетов сильно зависит от равномерности попадания данных в обе подгруппы. Так например я получил данные

Коэффициент точности : 0.55

test size = 0.10

Коэффициент точности : 0.63

test size = 0.20

Коэффициент точности : 0.61

test size = 0.30

Коэффициент точности : 0.55

test size = 0.40

Коэффициент точности : 0.52

test size = 0.50

Коэффициент точности : 0.42

test size = 0.60

Коэффициент точности : 0.35

test size = 0.70

Коэффициент точности : 0.27

test size = 0.80

При распределении фотографий в равномерном количестве качество расчетов существенно меняется