

Лабораторная работа №6. Оценка качества работы алгоритмов классификации и кластеризации.

1. На основе предыдущих лабораторных работ написать программу, оценивающую результаты работы различных алгоритмов кластеризации и классификации (k-means, kNN, SVM, максимин).
2. Необходимо реализовать следующие методы оценки:
 - а) Для внешнего оценивания:
 1. Точность, отклик (полнота), F1-мера
 2. ROC-кривые
 3. Среднеквадратичная ошибка
 4. Кросс-валидация
 5. Матрица неточностей (confusion matrix)
 - б) Для внутреннего оценивания:
 1. Коэффициент детерминации (R-squared для случая без известного разбиения)
 2. Коэффициент силуэта
 3. Индекс Данна (Dunn index)
 4. Индекс Девиса-Болдина (DBI)
 5. Индекс валидности по плотности кластеризации (DBCv)
3. Продемонстрировать работу программы на подготовленном наборе данных.
4. Функции, реализующие указанные способы оценки качества, должны быть реализованы студентом **самостоятельно**. Допускается использование готовых примитивов из стандартных или сторонних библиотек для выполнения базовых (в т.ч. векторизованных) операций (сложение, умножение, возведение в степень, вычисление среднего значения и т.п.).
5. Требования к набору данных: не менее 30 объектов, не менее 2 и не более 3 классов, не менее 3 и не более 7 атрибутов у объекта. Можно взять предложенные наборы данных «Ирисы» либо «Фрукты и овощи».
6. При выполнении задания рекомендуется использовать язык программирования Python или среду Octave.