



Машинное обучение

НИЯУ МИФИ, Кафедра финансового мониторинга

Лабораторный практикум.

В.Ю. Радыгин, Д.Ю. Куприянов, Т.А. Манаенкова

Семестр 1. Лабораторная работа 3

Лабораторная работа 3 рассчитана на два занятия и работу дома. Её целью является изучение основ классификации данных с помощью метода опорных векторов, метода главных компонент и расчёта характеристик качества классификатора.

Задание

1. Загрузите с сайта <https://sci2s.ugr.es/keel/datasets.php> набор статистических данных, указанный в вашем варианте. Разберитесь, какие данные приведены в наборе и какой атрибут является меткой класса.
2. На основе загруженного файла создайте Pandas DataFrame, подобрав правильные типы данных столбцов.
3. Выполните стандартизацию полученного дата фрейма.
4. Разделите дата фрейм на обучающую и тестовую выборки в соотношении 6 / 4.
5. На основе обучающей и тестовой выборки постройте SVM-классификатор, рассчитайте параметры эффективности классификатора (Accuracy). Меняя значение параметров kernel, gamma, coef0, degree, C (на основе вариантов, представленных в лекции) обосновано подберите наиболее удачное решение классификации для подготовленных выборок.
6. Применить метод главных компонент. Заново построить svm-классификатор с выбранным кол-вом классов. Проанализировать результаты.
7. **Необязательное.** Посторить предиктивный эллипс, проанализировать полученные результаты.

Примечания:

- Параметры: gamma, coef0, degree, C — лучше брать разумные числа (проанализировать, какие более подходящие).

Варианты

1. <https://sci2s.ugr.es/keel/dataset.php?cod=210>
2. <https://sci2s.ugr.es/keel/dataset.php?cod=209>
3. <https://sci2s.ugr.es/keel/dataset.php?cod=107>
4. <https://sci2s.ugr.es/keel/dataset.php?cod=72>

Новые:

1. <https://www.kaggle.com/datasets/chandavandann/nasa-exoplanet-ml-superdataset>
2. <https://www.kaggle.com/datasets/ehsanemesmaeli/red-and-white-wine-quality-merged>
3. Тут можно придумать свои 😊