



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение высшего

образования

«Дальневосточный федеральный университет»

(ДВФУ)

ИНСТИТУТ МАТЕМАТИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ

(ШКОЛА)

Департамент информационных и компьютерных систем

ОТЧЕТ

по дисциплине «системы искусственного интеллекта»

Выполнил студенты группы Б9122-
09.03.03пикд

Зверев Р. И.

Проверил преподаватель

Бочарова В. В.

зачтено/не зачтено

г. Владивосток

2025 г

Оглавление

Цель работы	3
Введение	4
Описание данных.....	5
Сравнение SVM	6
Нахождение лучших параметров SVC	7
Подбор других параметров SVC	8
Заключение.....	9
Список литературы.....	10

Цель работы

Целью работы является применение метода опорных векторов, SVM, для задачи бинарной классификации для определения фальшивых купюр в датасете BanknoteAuthentication.

Постановка задачи

В данной работе рассматривается задача классификации купюр на настоящие и фальшивые на примере датасета Banknote Authentication.

Необходимо реализовать следующие этапы и функции:

- Рассмотреть реализации SVM в разных библиотеках;
- Подобрать лучшие параметры (kernel, max_iter) для SVM в sklearn;
- Подобрать лучшие параметры SVM, используя весь набор параметров (C, gamma и другие).

Введение

Подделка банкнот — практическая задача с серьёзными экономическими последствиями; автоматическое отделение настоящих купюр от фальшивых требует надёжных методов классификации. В этой лабораторной работе мы используем метод опорных векторов (SVM), поскольку он хорошо подходит для задач бинарной классификации благодаря жёсткой границе разделения, возможностям ядерных преобразований и контролю за регуляризацией.

На конкретном примере датасета Banknote Authentication необходимо построить и сравнить SVM-модели, подобрать ядро и параметры обучения, а также исследовать влияние масштабирования и выбора гиперпараметров на качество классификации. В результате вы научитесь правильно настраивать SVM, интерпретировать метрики качества, и делать выводы о практической применимости модели для обнаружения фальшивых купюр.

Описание данных

Описание набора данных Banknote Authentication.

Для проведения анализа был выбран набор данных Banknote Authentication, который содержит данные о купюрах. Датасет включает 1372 записи и 5 признаков(столбцов).

Полное описание признаков:

- **Variance:** коэффициенты дисперсии;
- **Skewness:** коэффициенты асимметрии;
- **Curtosis:** эксцесса вейвлет-преобразования изображения;
- **Entropy:** энтропия изображения;
- **Class:** класс купюры, 1 — фальшивка, 0 — настоящая.

Сравнение SVM

В качестве моделей SVM для сравнения были взяты: SVC (support vector classifier) из sklearn и LinearSVC из PySpark.

Для модели SVC были заданы параметры: kernel = «rbf», max_iter = 200000, cache_size = 8000, остальные параметры заданы библиотекой.

Для модели LinearSVC были заданы параметры: maxIter = 200000, maxBlockSizeInMB = 8000, остальные параметры были заданы библиотекой.

В результате SVC выдал результат в 99,56% accuracy, а LinearSVC выдал 99,115%.

SVC оказался лучше, так как, в отличие от PySpark конкурента, предоставляет не только линейную реализацию SVM, но еще и нелинейную с RBF ядром.

Нахождение лучших параметров SVC

Найдем лучшую комбинацию из `kernel` и `max_iter` для SVC модели с помощью `GridSearch`. Переберем `kernel` из набора библиотеки: «linear», «rbf», «poly», «sigmoid». Переберем `max_iter` из заданных значений: -1, 200000, 400000.

В итоге лучшая пара комбинаций: `max_iter = -1`, `kernel = rbf`. Такая пара выдает 99,65% точности.

Подбор других параметров SVC

Проверим можно ли улучшить точность классификации, подобрав не только `kernel` и `max_iter`, но и другие параметры.

Добавим в набор поиска `GridSearch` еще параметры:

- `C`: 0.5, 1.5, 2, 2.5;
- `Gamma`: scale, auto, 1.0, 1.5;
- `Degree`: 2, 3, 4;
- `Class_weight`: None, balanced, {0: 762/1372, 1: 610/1372}.

В итоге лучший набор параметров: `C = 0.5`, `class_weight = None`, `degree = 2`, `gamma = auto`, `kernel = rbf`, `max_iter = -1`. SVC с такими параметрами выдает точность 100% на датасете.

Заключение

В работе на датасете Banknote Authentication сравнивались SVC и LinearSVC.

LinearSVC показал точность ~99.12%, SVC же с RBF ядром показал 99.56 — 99.65%, а после подбора полного набора параметров показал точность 100%.

Лучшая модель — SVC с параметрами: $C = 0.5$, `class_weight = None`, `degree = 2`, `gamma = auto`, `kernel = rbf`, `max_iter = -1`, достигающая 100% accuracy на данном датасете.

Список литературы

1. GitHub: исходный код лабораторной работы. – URL: [Лабораторная работа №3.5](#) (дата обращения: [09.10.2025]). – Текст: электронный.