

Рецензия на статью «Автоматический подбор гиперпараметров моделей машинного обучения» (Надршена В.Р., Китов В.В.)

Рецензент: Мельник Юрий

1 декабря 2023 г.

В статье производится сравнение различных методов оптимизации гиперпараметров, таких как случайный поиск (в частности его реализация `RandomizedSearchCV` из `scikit-learn`), направленный поиск из фреймворка `Optuna` и байесовская оптимизация гиперпараметров (`BayesSearchCV` из `scikit-optimize`) на примере решения задачи бинарной классификации с помощью такой классической модели машинного обучения, как `SVM Classifier`. В качестве набора данных использовался датасет `Wisconsin Diagnostic Breast Cancer (WDBC)`, который содержит диагностические измерения рака груди, каждый экземпляр представляет собой оцифрованное изображение мелкоигольной аспирационной биопсии (FNA) опухоли груди. Набор данных предоставляет характеристики, вычисленные на основе этих изображений, и использует их для классификации наблюдений как доброкачественные или злокачественные.

В рамках данной работы были получены интересные экспериментальные результаты по сравнению рассматриваемых методов оптимизации как с точки зрения метрик качества обученных при найденных значениях гиперпараметров моделей, так и в смысле времени, затраченного на подбор конкретных значений гиперпараметров. На основании полученных результатов были сделаны выводы насчёт уместности использования каждого из рассматриваемых алгоритмов оптимизации гиперпараметров при решении данной задачи классификации.

Замечания:

- Перед описанием результатов применения методов оптимизации стоит указать, какие конкретно гиперпараметры есть у выбранной модели метода опорных векторов, по каким из них велась оптимизация.
- При визуализации результатов сравнения точности моделей в заголовке графика стоит указать, что оно производилось на тестовой выборке (возможно, это очевидно, но всё же у читателя могут возникнуть некоторые вопросы на этот счёт).
- В выводах стоит указать, какой метод оптимизации гиперпараметров показал наилучшие результаты по соотношению точность-время при решении выбранной задачи классификации.