

Выполнил(а) Абдулов И.А., № группы 3121, дата 03.12.2022, оценка

Название статьи/главы книги: Исследование оптимизации гиперпараметров алгоритма k-ближайших соседей		
ФИО авторов статьи: Трясучкин В.А., Синцева М.М.	Дата публикации: 2019 г.	Размер статьи 6 стр.
Прямая полная ссылка на источник и сокращенная ссылка: https://cyberleninka.ru/article/n/razrabotka-veb-prilozheniya-dlya-obrabotki-i-predstavleniya-prostranstvennyh-metadannyh-geoportala http://bit.ly/3XjW6E8		
Тэги, ключевые слова или словосочетания Машинное обучение, оптимизация, классификация, алгоритмы, k-ближайших соседей		
Перечень фактов, упомянутых в статье: Машинное обучение активно применяется в жизни людей. Классификация является разделом машинного обучения. Обучающая выборка в машинном обучении – конечное множество объектов, для которых определена принадлежность к классу. Для остальных объектов классы не определены. В данном исследовании у авторов появляется необходимость создать алгоритм, определяющий класс объекта. Алгоритм k-ближайших соседей относится к алгоритмам классификации, относит объект к тому классу, которому принадлежит k-ближайших к нему объектов из обучающей выборки, k – заданное число. То есть для данного алгоритма определена определенная математическая формула. При решении задач классификации каких-либо объектов методом k-ближайших соседей перед разработчиком ставится задача автоматизации выбора числового параметра k, ведь ручной подбор этого гиперпараметра занимает длительное время и даже после подбора, не всегда удастся добиться максимальной точности. В связи с данной сложностью исследование методов оптимизации гиперпараметров является актуальной задачей. Существует несколько методов оптимизации: gridsearch и randomsearch, которые позволяют повысить скорость и точность решения задачи. Реализация данных методов оптимизации позволила ускорить подбор гиперпараметра. В качестве особенностей нужно выделить, что gridsearch стоит применять только в том случае, когда имеется небольшой объем данных для обработки или требуется максимальная точность. В случае, когда в задаче на первом месте стоит скорость обработки данных, то стоит использовать randomsearch.		
Позитивные следствия и/или достоинства описанной в статье технологии -Оптимизация gridsearch наиболее точна -Python пользуется популярностью среди разработчиков в области машинного обучения благодаря своей простоте и наличию большого количества библиотек -Для визуализации графика с данными используется библиотека matplotlib		
Негативные следствия и/или недостатки описанной в статье технологии -Метод gridsearch зависит от вычислительной мощности компьютера -Метод randomsearch является менее точным -В задаче классификации не обойтись без обучающей выборки		
Ваши замечания, пожелания преподавателю или анекдот о программистах¹		

¹ – не ведёт к повышению или понижению оценки