

Домашнее задание к лекции «Задачи классификации и кластеризации»



Задание

[Вопросы по заданию](#)

Преподаватель: Олег Булыгин

Домашнее задание

<http://prntscr.com/umci7o>

1. Возьмите датасет с цветками iris'a (функция load_iris из библиотеки sklearn)
2. Оставьте два признака - sepal_length и sepal_width и целевую переменную - variety
3. Разделите данные на выборку для обучения и тестирования
4. Постройте модель LDA
5. Визуализируйте предсказания для тестовой выборки и центры классов
6. Отбросьте целевую переменную и оставьте только два признака - sepal_length и sepal_width
7. Подберите оптимальное число кластеров для алгоритма kmeans и визуализируйте полученную кластеризацию

Решение

Зачет

https://github.com/PeterM-lab/PYDA1/blob/main/17_LDA_cluster/Untitled.ipynb 10 мая 2021

LINK

Вы загрузили решение 10 мая в 21:11

Олег Булыгин поставил(а) зачет 14 мая в 09:39



Олег Булыгин

ПРЕПОДАВАТЕЛЬ

14 мая 2021 09:39

Петр, приветствую!

Спасибо, все супер.

Как вы могли заметить, результат может очень сильно отличаться в зависимости от случайного разбиения данных на train и test. Такое может происходить, особенно когда данных не очень много. В реальности просто так это оставлять нельзя (иначе мы не сможем объективно сравнивать разные модели между собой), всегда применяются различные механизмы кросс-валидации (то есть при обучении модели данные разбиваются множеством разных способов и считается условно "усредненный" лучший результат). Рекомендую ознакомиться вот с этим примером, там на этих же данных применяется этот механизм в самой своей простой форме: www.statology.org...in-python/

Зачет!



Вам понравилось?

Вопросы по заданию

Задайте вопрос — вам помогут одногруппники и эксперты

Добавьте более подробное описание вашего вопроса, если необходимо

Спросить

Все вопросы Помочь с ответом

Пока нет вопросов