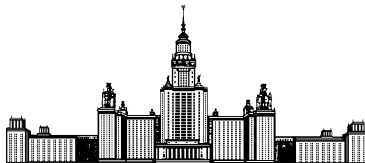


Московский государственный университет имени М. В. Ломоносова



Факультет Вычислительной Математики и Кибернетики

Кафедра Математических Методов Прогнозирования

КУРСОВАЯ РАБОТА СТУДЕНТА 317 ГРУППЫ

«Название курсовой работы»

Выполнил:

студент 3 курса 317 группы

Каюмов Эмиль Марселевич

Научный руководитель:

д.ф.-м.н., профессор

Дьяконов Александр Геннадьевич

Содержание

1	Введение	2
2	Методы восстановления пропусков	2
2.1	Игнорирование объектов с пропущенными значениями	2
2.2	Замена специальным значением	2
2.3	Замена самым частым или средним значением	2
2.4	Замена с помощью SVD	2
2.5	Замена с помощью метода ближайших соседей	2
2.6	Замена с помощью случайного леса	2
2.7	Замена с помощью линейной регрессии	2
2.8	Замена с помощью EM-алгорима	2
2.9	Замена с помощью метода k средних	2
3	Эксперименты с данными	2
3.1	Исходные данные и условия эксперимента	2
3.2	Результаты эксперимента	3
3.3	Обсуждение и выводы	3
4	Заключение	3
	Список литературы	3

1 Введение

О том, как это часто встречается и почему это может быть нужно.

Если написать классификацию MCAR, MAR, NMAR, то это, наверное, должно быть здесь. Но тогда надо объяснять, какой метод на что рассчитан (что сомнительно в какой-то мере).

2 Методы восстановления пропусков

Возможно, стоит делить на группы: простейшие методы (первые три), основанные на предсказаниях (логистическая регрессия, случайный лес, ...), основанные на разложениях и подобных вещах (SVD, EM, K-Means, ...).

Здесь же про постановку задачи, обозначения.

2.1 Игнорирование объектов с пропущенными значениями

2.2 Замена специальным значением

2.3 Замена самым частым или средним значением

2.4 Замена с помощью SVD

2.5 Замена с помощью метода ближайших соседей

2.6 Замена с помощью случайного леса

2.7 Замена с помощью линейной регрессии

2.8 Замена с помощью EM-алгорима

2.9 Замена с помощью метода k средних

3 Эксперименты с данными

3.1 Исходные данные и условия эксперимента

Описание трех датасетов с полностью заполненными значениями. Описание методов удаления данных.

Описание трех датасетов с натуральными пропусками в данных.

Описание того, как проводятся измерения.

3.2 Результаты эксперимента

Графики и таблицы с результатами.

3.3 Обсуждение и выводы

В каких случаях что использовать. Что лучше работает, что не стоит использовать.

4 Заключение

Список литературы