

Задание: Моделирование пропущенных данных

Задана таблица, содержащая 2701 строку и 128 столбцов. Элементы таблицы – это наблюдаемые индикаторы коммерческой активности 2701 магазинов по отношению к 128 типам товаров. Не все активности наблюдаются, ненаблюдаемые значения заменены 0. Присутствующие ненулевые наблюдения содержат статистические шумы.

Требуется установить наличие информационной избыточности в измерениях и выполнить прогноз пропущенных значений на основе совокупности имеющихся. Нужно также оценить точность такого прогноза. Может ли построенная статистическая модель исправить возможные искажения в присутствующих данных?

Опишите Ваше решение, и приложите, пожалуйста, тексты программ. Если были использованы внешние компоненты и библиотеки, то укажите их. Предпочтителен язык R.

Приложение: файл task3.txt