Министерство образования Республики Беларусь

Учреждение образования

БелорусскиЙ государственный университет

информатики и радиоэлектроники

Факультет компьютерных систем и сетей

Кафедра информатики

ОТЧЕТ

по лабораторной работе №8

**Выявление аномалий**

|  |  |  |
| --- | --- | --- |
| Студент |  | И.Ю. Фалько |
| Преподаватель |  | М. В. Стержанов |

Минск 2019

СОДЕРЖАНИЕ

[Условие лабораторной работы 3](#_Toc25747475)

[Ход выполнения 3](#_Toc25747476)

[Заключение 7](#_Toc25747477)

# Условие лабораторной работы

Набор данных **ex8data1.mat** представляет собой файл формата \*.mat (т.е. сохраненного из Matlab). Набор содержит две переменные X1 и X2 - задержка в мс и пропускная способность в мб/c серверов. Среди серверов необходимо выделить те, характеристики которых аномальные. Набор разделен на обучающую выборку (X), которая не содержит меток классов, а также валидационную (Xval, yval), на которой необходимо оценить качество алгоритма выявления аномалий. В метках классов 0 обозначает отсутствие аномалии, а 1, соответственно, ее наличие.

Набор данных **ex8data2.mat** представляет собой файл формата \*.mat (т.е. сохраненного из Matlab). Набор содержит 11-мерную переменную X - координаты точек, среди которых необходимо выделить аномальные. Набор разделен на обучающую выборку (X), которая не содержит меток классов, а также валидационную (Xval, yval), на которой необходимо оценить качество алгоритма выявления аномалий.

# Ход выполнения

1. Загрузите данные **ex8data1.mat** из файла.
2. Постройте график загруженных данных в виде диаграммы рассеяния.

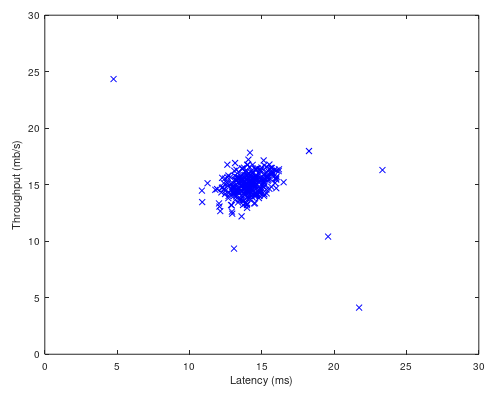
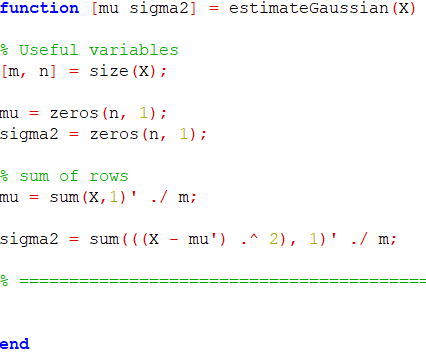


Рисунок 1 – График загруженных данных (ex8data1.mat)

3-5 Представьте данные в виде двух независимых нормально распределенных случайных величин.

Оцените параметры распределений случайных величин.

Постройте график плотности распределения получившейся случайной величины в виде изолиний, совместив его с графиком из пункта 2.



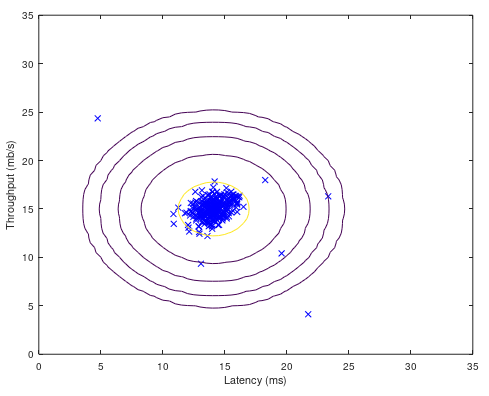
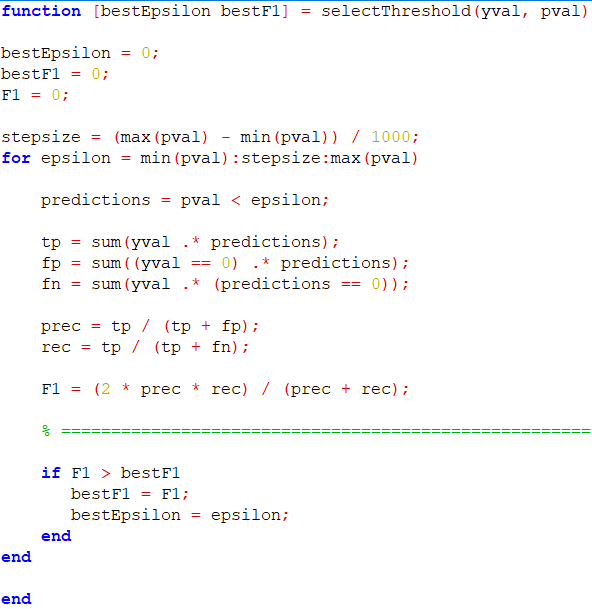


Рисунок 2 – График плотности распределения получившейся случайной величины в виде изолиний совместно с исходными данными

1. Подберите значение порога для обнаружения аномалий на основе валидационной выборки. В качестве метрики используйте F1-меру.





1. Выделите аномальные наблюдения на графике из пункта 5 с учетом выбранного порогового значения.

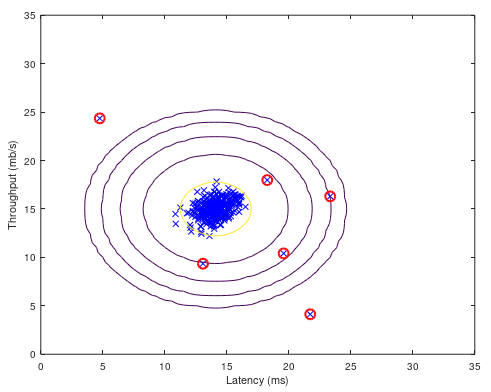


Рисунок 3 – График с выделением аномальных наблюдений

* 1. Загрузите данные **ex8data2.mat** из файла.

Представьте данные в виде 11-мерной нормально распределенной случайной величины.

Оцените параметры распределения случайной величины.

Подберите значение порога для обнаружения аномалий на основе валидационной выборки. В качестве метрики используйте F1-меру.

1. Выделите аномальные наблюдения в обучающей выборке.

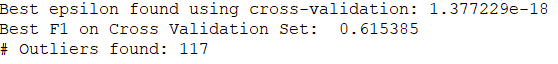


Рисунок 4 – Результаты подбора значения порога для обнаружения аномалий для данных из файла ex8data2.mat

# ЗАКЛЮЧЕНИЕ

В данной лабораторной работе было освоено выявление аномалии с помощью Гауссового распределения, подбор значения порога для обнаружения аномалий на основе валидационной выборки. Метод выявления аномалий достаточно похож на обучение с учителем, но они имеют разные области применения.