



MITx 6.86x

Machine Learning with Python-From Linear Models to Deep Learning

[Course](#)

[Progress](#)

[Dates](#)

[Discussion](#)

[Resources](#)



[Course](#) / [Unit 1. Linear Classifiers and Generalizatio...](#) / [Project 1: Automatic R](#)

< Previous



1. Introduction

 Bookmark this page

The goal of this project is to design a classifier to use for sentiment analysis of product reviews. The dataset consists of reviews written by Amazon customers for various food products. The reviews, originally on a 5 point scale, have been adjusted to a +1 or -1 scale, representing a positive or negative sentiment respectively.

Below are two example entries from our dataset. Each entry consists of the review and its sentiment label. The reviews were written by different customers describing their experience with a sugar-free candy.

Review
<i>Nasty No flavor. The candy is just red, No flavor. Just plan and chewy. I would never buy them again.</i>
<i>YUMMY! You would never guess that they're sugar-free and it's so great that you can eat them pretty much guilt free! i was so impressed that i've ordered some for myself (w dark chocolate) to take to the office. These are just EXCELLENT!</i>

In order to automatically analyze reviews, you will need to complete the following tasks:

1. Implement and compare three types of linear classifiers: the **perceptron** algorithm, the **logistic regression** algorithm, and the **Pegasos** algorithm.
2. Use your classifiers on the food review dataset, using some simple text features.
3. Experiment with additional features and explore their impact on classifier performance.

Setup Details:

For this project and throughout the course we will be using Python 3.8 with some additional libraries. We strongly recommend that you take note of how the NumPy numerical library is used in the course materials. Please read through the on-line NumPy tutorial. **NumPy arrays are much more efficient than lists when doing numerical computation. In addition, using NumPy will substantially reduce the amount of code you will need to write.**

1. *Note on software: For this project, you will need the **NumPy** numerical toolbox, and the **scikit-learn** machine learning toolbox.*
2. Download [sentiment_analysis.tar.gz](#) and untar it into a working directory. The sentiment_analysis directory contains the various data files in **.tsv** format, along with the following python files:

- **project1.py** contains various useful functions and function templates that you will use to implement the learning algorithms.

to fully check correctness and receive your grade for individual function implementation.

can also implement the function for the first part of the project.

project1.py file

[< Previous](#)[Next >](#)

second development flow.



edX

[About](#)[Affiliates](#)[edX for Business](#)[Open edX](#)[Careers](#)[News](#)

Legal

[Terms of Service & Honor Code](#)[Privacy Policy](#)[Accessibility Policy](#)[Trademark Policy](#)[Sitemap](#)[Cookie Policy](#)[Do Not Sell My Personal Information](#)

Connect

[Blog](#)[Contact Us](#)[Help Center](#)[Security](#)[Media Kit](#)