

به نام خدا



تاریخ: ۱۴۰۰/۱۰/۱۷

مهرانه مقتدائی فر ۹۷۲۲۲۰۸۶

استاد درس: دکتر خردپیشه

درس: مبانی علوم داده

گزارش تمرین سری ۴ - بخش اول

در این بخش به پیاده سازی مدل LinearRegression پرداخته شده است. با توجه به فرمول های داده شده در صورت تمرین، توابع مورد نظر پیاده سازی شده است.

این توابع به شرح زیر است:

تابع Gradient_Decent:

با توجه به الگوریتم مربوطه، مقدار θ در هر مرحله آپدیت میشود.

تابع Compute Cost:

در این تابع مقدار loss function محاسبه میشود، که این مقدار با توجه به X , Y در هر مرحله متفاوت خواهد بود و با توجه به میزان θ که از الگوریتم gradient decent بدست می آید این مقدار تغییر میکند و در هر iteration باید کاهش پیدا کند.

در این تابع مقدار MSE محاسبه میشود.

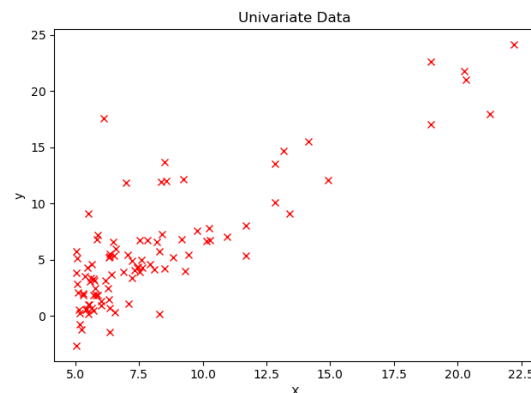
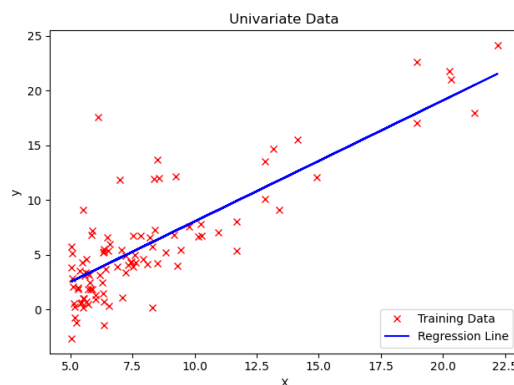
تابع predict:

در این تابع نیز تنها کافیست مقدار predict شده را که مدل با توجه به ورودی ها پیشبینی میکند، برگردانیم

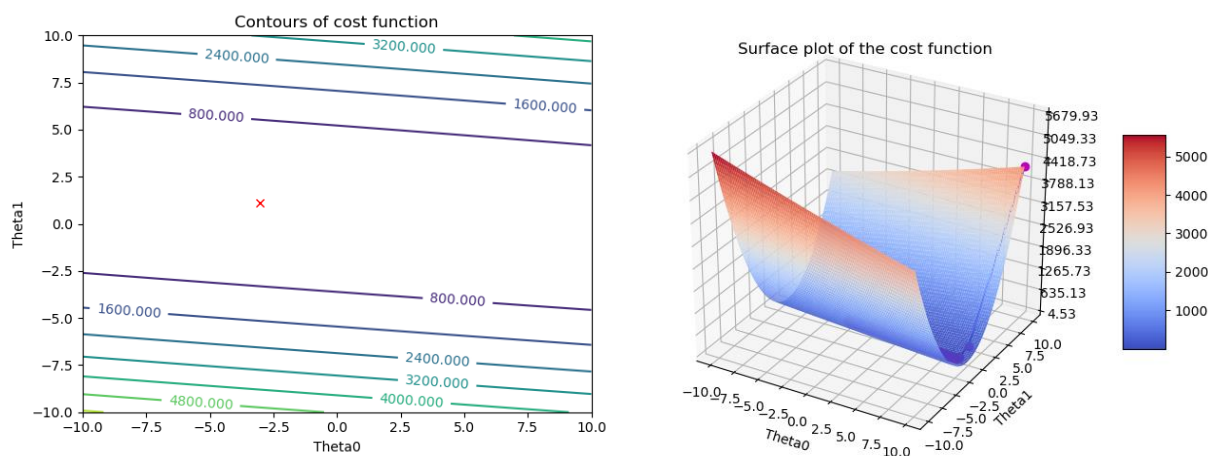
حال نتایج دو مدل univariate و multivariate را بررسی میکنیم:

• UNIVARIATE

در حالت تک متغیره، بر روی دیتاست univariate.dat مدل را تست کردیم. همانطور که میبینیم، نمودار های زیر نشان دهنده مدل regression بر روی داده ها می باشد.



نمودار های مربوط به cost function را نیز در زیر میبینیم:



همچنین با توجه به این مدل، مقدار خطای MSE پس از 1500 تکرار برابر با 4.544 است.

MULTIVARIATE •

در این حالت، مدل `linearRegression` را بر روی دیتاست `multivariate.dat` اعمال کردیم. سپس مقدار `cost` را نیز برای این مدل محاسبه کردیم. پس از 2000 تکرار مقدار MSE حدود 20×10^8 است. با توجه به آنکه این دیتاست دارای داده های چند متغیره است، مقدار `loss` زیاد خواهد شد.

HOLDOUT_TEST •

در این دیتاست، تعدادی داده تست داده شده است. این داده ها مقادیر چند متغیره هستند و باید مانند دیتاست های پیشین، مدل رگرسیون چندمتغیره را بر روی این دیتاست نیز اعمال کنیم. از آنجایی که خواسته شده مقدار `RMSE` را گزارش کنیم، تنها نیاز است که مقدار جذر `mean_squared_error` را محاسبه کنیم. پس از 2000 تکرار، مقدار `RMSE` برای این دیتا ها تقریباً حدود 32000 تا شده است.