|  |  |  |
| --- | --- | --- |
|  | به نام خدا |  |
| **دانشگاه تهران**  **دانشکده‌ی مهندسی برق و کامپیوتر**  **شبکه عصبی**  **گزارش پروژه ۳** | | |

|  |  |  |
| --- | --- | --- |
| ملیکه احقاقی | سید علی طباطبایی آل طه | نام و نام خانوادگی |
| ۸۱۰۱۹۴۲۵۴ | ۸۱۰۱۹۴۴۶۲ | شماره‌ی دانشجویی |
| ۲۳/۳/۹۷ | | تاریخ ارسال گزارش |

­

* 1. **چکیده**

این پروژه از سه فاز رگرسیون، Q-learning و خوشه بندی تشکیل شده است. در فاز اول در پیاده سازی بخش رگرسیون تک متغیره و چندمتغیره هدف به دست آوردن یک رابطه ی خطی بین خروجی و ویژگی های اراءه شده در داده های ورودی است. در ادامه به بررسی logistic regression پرداختیم که در این بخش هدف پیدا کردن یک مدل طبقه بند بین دو دسته داده است که بهترین خط که مرز بین این دو دسته داده را مشخص می کند پیدا شود. روش بعدی دسته بندی که در پیش گرفته شد SVM بود که در آن نیز هدف دسته بندی داده هاست با این تفاوت که بسته به کارگیری کرنل های گوناگون قادر به یافتن مرزها ی خطی، گاوسی ، چند جمله ای و ... هستیم و پیاده سازی متفاوتی خواهد داشت.

در فاز Q-learning به دنبال حل یک ماز در یک محیط unknown است. برای حل این مساله باید عامل هوشمند به مرور زمان محیط خود را بشناسد و با توجه به reward function ای که توسط ما مشخص می شود بهترین مسیر خود به سمت هدف را پیدا کند و به مرور knowledge خود نسبت به محیط و ماتریس <state,action> خود را کامل کند.  
و در نهایت در مرحله ی خوشه بندی به دنبال گروه بندی داده ها هستیم که با اختیار کردن مراکز تصادفی برای داده ها بتوان به شکل بهینه ای به حداقل تعداد ممکن داده ها را دسته بندی کرد و با انتخاب بهترین مراکز فواصل داده های یک گروه نسبت به مرکز گروه را کاهش داد. برای پیاده سازی clustering از دو الگوریتم k-means وFuzzy C-means استفاده شد.

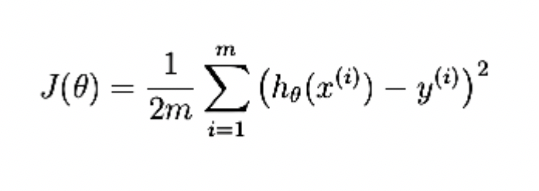
* 1. **ارائه‌ی روش**

**۱. رگرسیون:**

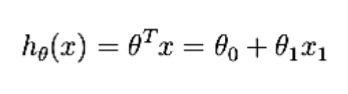
**۱.۱. رگرسیون خطی:**

**۱.۱.۱. رگرسیون خطی تک متغیره:**

در این بخش برای انجام پیش بینی، از یکی از ویژگی های مربوطه استفاده میکنیم. هدف کمینه کردن تابع هزینه که آن را تابع MSE در نظرگرفته ایم، است.



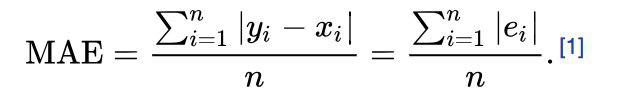
که در مدل خطی داریم:



برای بهینه سازی و کمینه کردن این تابع و ضرایب θ از روش گرادیان کاهشی استفاده میکنیم تا جایی که θ بهینه شده و تابع هزینه به مقدار کمینه خود برسد.

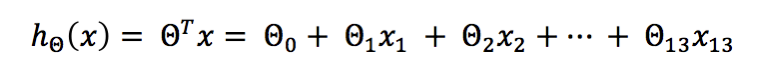
با توجه به خواسته صورت پروژه در این بخش از ویژگی های اول و سوم داده هایمان استفاده می کنیم.

برای محاسبه خطای میانگین تمام نمونه ها از فرمول زیر استفاده می کنیم:



**۱.۱.۲. رگرسیون خطی چند متغیره:**

در این بخش با توجه به تمام ویژگی ها، قیمت مربوطه را پیش بینی می کنیم. تابع خطا نیز به صورت قسمت قبل است و تنها تفاوت در تابع زیر می باشد:



* 1. **ارائه‌ی نتایج**

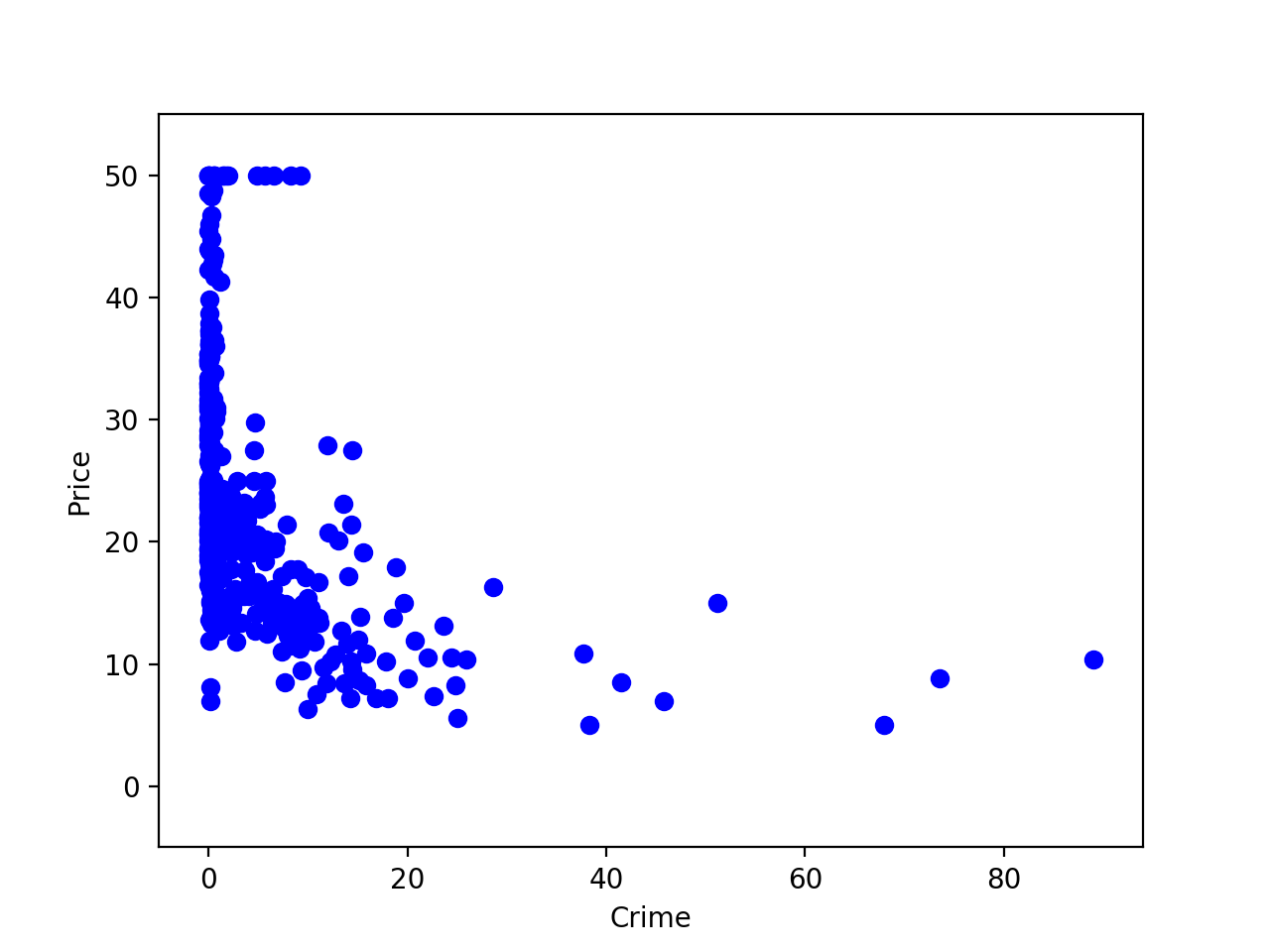
**۱. رگرسیون:**

**۱.۱. رگرسیون خطی:**

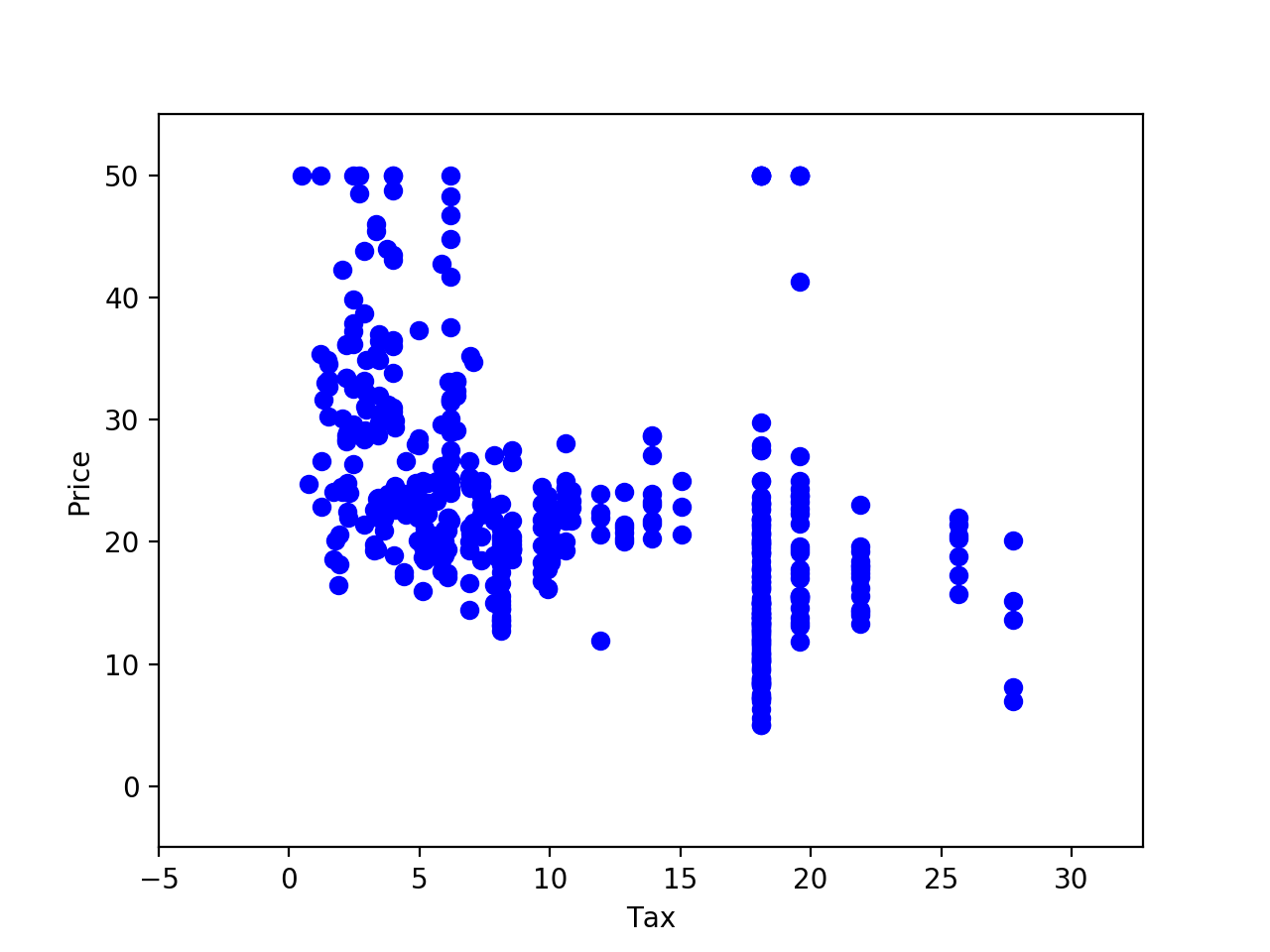
**۱.۱.۱. رگرسیون تک متغیره:**

نمودار قیمت خانه برحسب هر یک از دو ویژگی:

۱- قیمت خانه برحسب تعداد جرایم انجام شده در منطقه



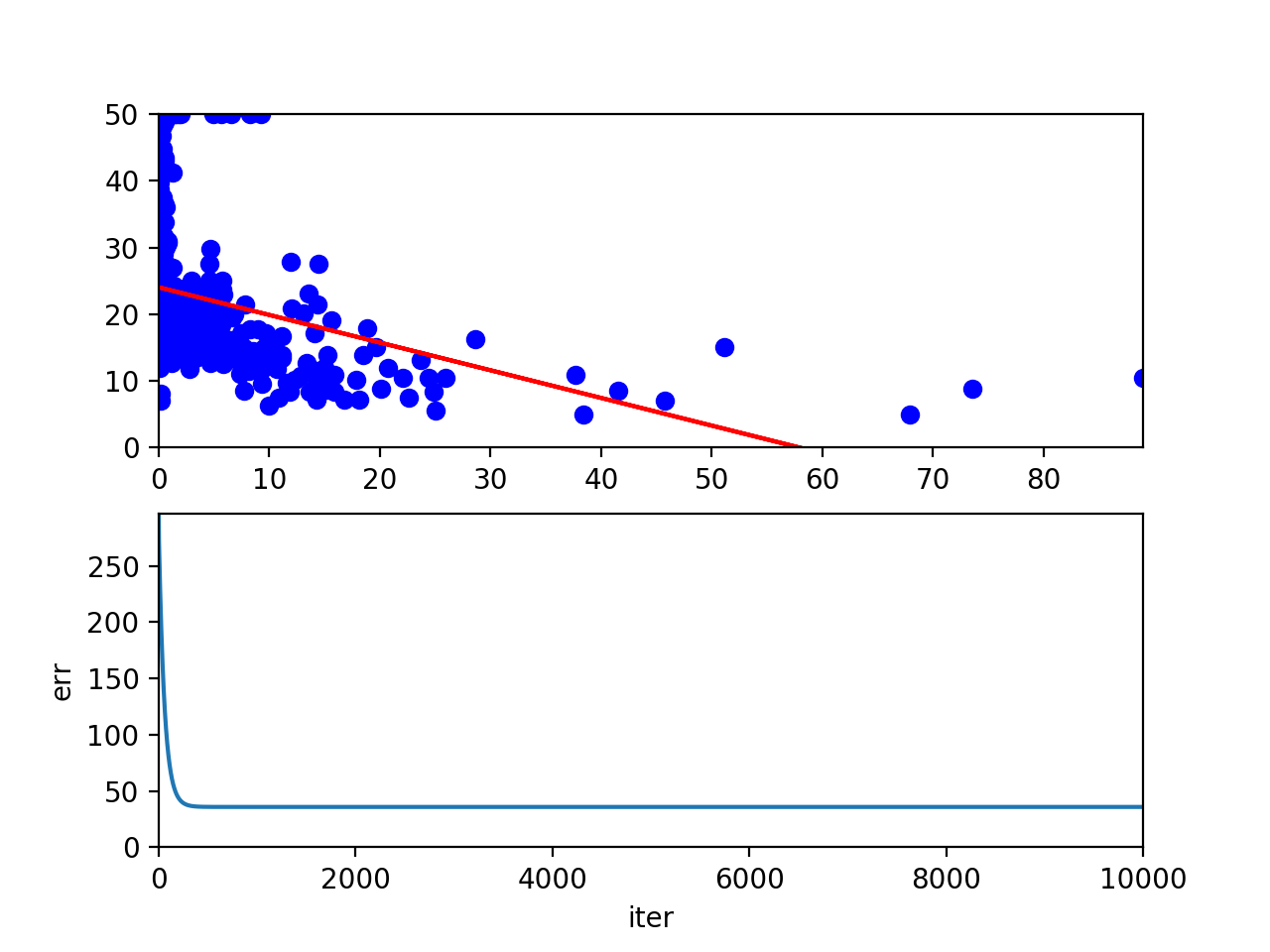
۲- قیمت خانه برحسب مقدار مالیات ملک



خط محاسبه شده و تغییرات تابع هزینه در هر iteration:

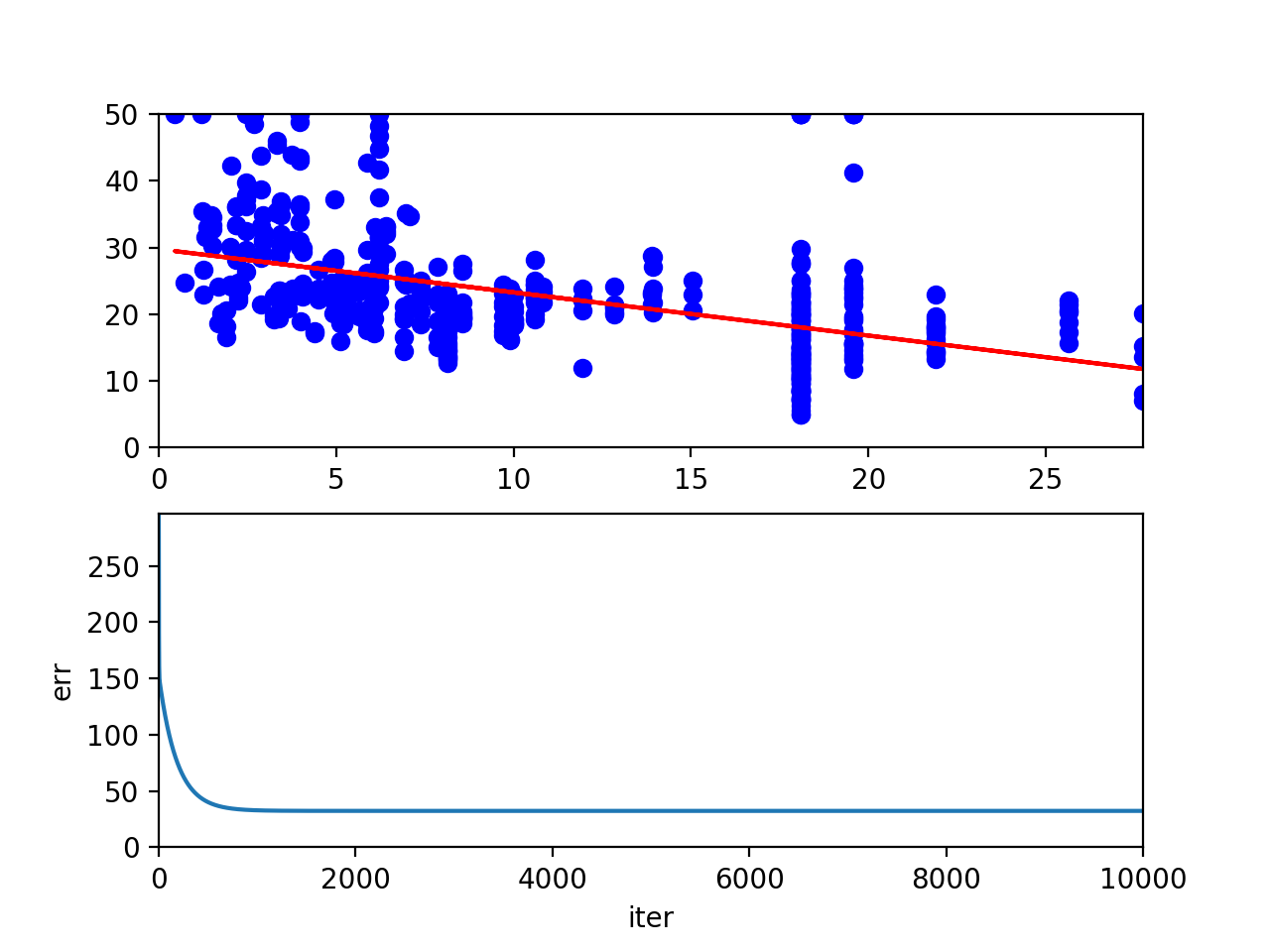
۱- ویژگی تعداد جرایم انجام شده در منطقه

y = -0.415190x + 24.033106



۲- ویژگی مقدار مالیات ملک

y = -0.648490x + 29.754897



مقدار خطای میانگین:

۱- ویژگی تعداد جرایم انجام شده در منطقه

MSE = 35.845368 تابع هزینه بعد از آخرین تکرار

MAE = 6.135297 مقدار خطای میانگین

۲- ویژگی مقدار مالیات ملک

MSE = 32.333111 تابع هزینه بعد از آخرین تکرار

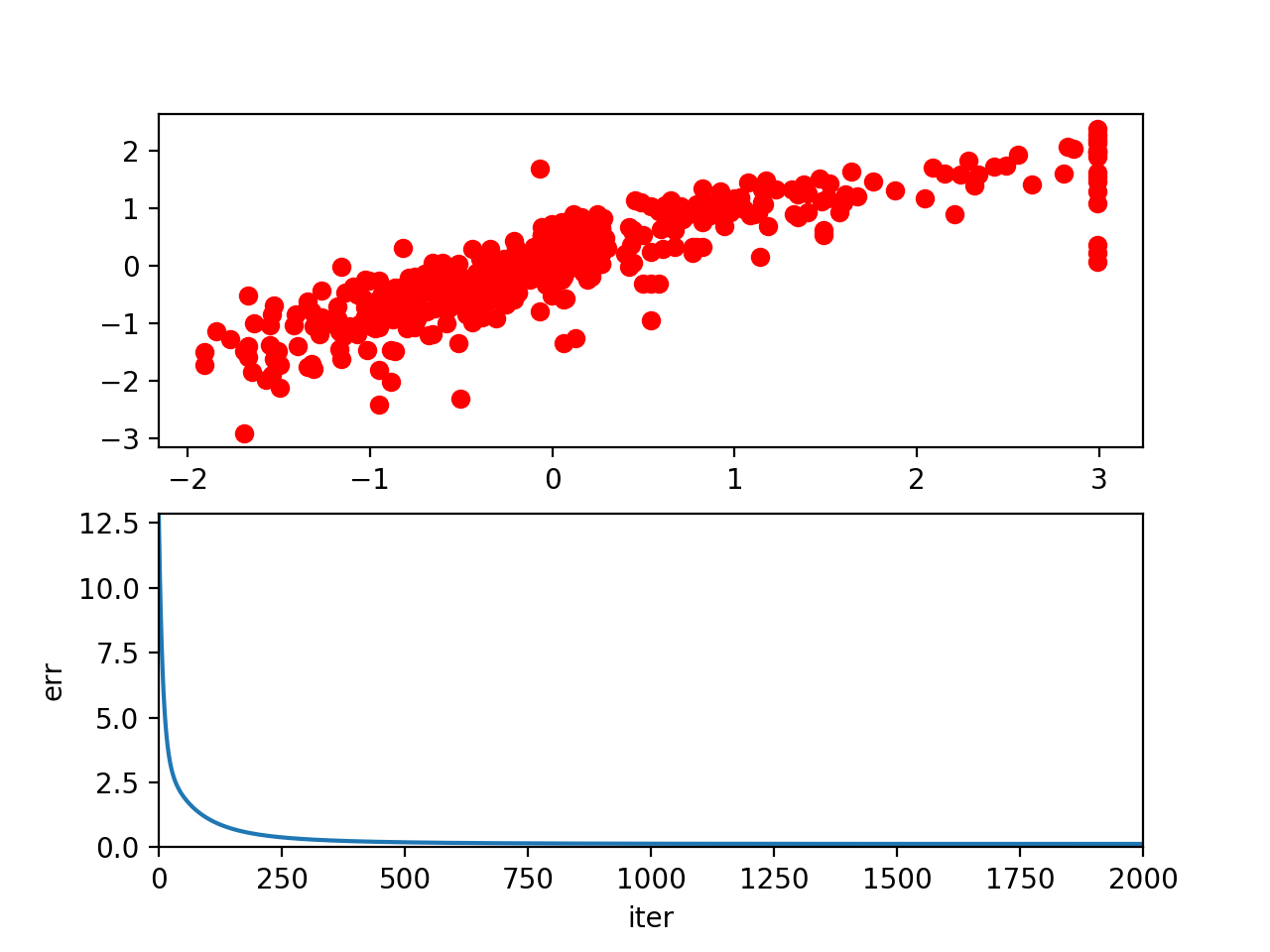
MAE = 5.776538 مقدار خطای میانگین

* مقادیر بالا در ۱۰۰۰۰ تکرار و با learning rate برابر با ۰.۰۱ بدست آمده اند که به صورت تجربی و با آزمون و خطا بدست آمده اند.

**۱.۱.۱. رگرسیون چند متغیره:**

در ابتدا بدون استفاده از روش l2norm بررسی میکنیم و نتایج را ارائه می دهیم:

num\_of\_iter = 2000  
learning\_rate = 0.01



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| θ13 | θ12 | θ11 | θ10 | θ9 | θ8 | θ7 | θ6 | θ5 | θ4 | θ3 | θ2 | θ1 | θ0 |
| 1.863e-09 | -9.097e-02 | 1.008e-01 | 1.180e-03 | 7.659e-02 | -1.882e-01 | 3.048e-01 | -3.972e-03 | -3.052e-01 | 2.332e-01 | -1.770e-01 | -2.173e-01 | 9.441e-02 | -4.031e-01 |

MSE = 0.13010352617776446693

MAE = 0.35354468514984704886

با استفاده از روش L2 norm:

* 1. **تحلیل نتایج**

**۱. رگرسیون:**

**۱.۱. رگرسیون خطی:**

**۱.۱.۱. رگرسیون تک متغیره:**

اگر تعداد تکرارها کم باشد ممکن است به بهترین جواب ممکن نرسیم و اگر learning rate کمتر باشد ممکن است در مینیمم محلی گیر بیفتد و اگر بزرگتر باشد ممکن است مینیمم گلوبال را از دست بدهد

* 1. **جمع‌بندی و نتیجه‌گیری**