测试集的结果真实值为

```
[[19.3 23.1 19.3 14.6 16.  22.1 35.2 28.4 15.7 17.4 20.6 12.1  7.2 24.6
  22.  17.9 22.2 20.1 22.7 15.6 10.8 19.9 21.4 17.8 23.2 24.5 22.5 16.2
  19.8  8.3  8.7  7.4 23.1  6.3 23.  16.4  7.  16.8 34.9 22.6 25.  35.1
  50.  21.9 21.8 35.4 33.3 21.2  9.5 37.6  7.5]]
```

使用解析法求得的系数如下：

```
[[-6.30121603e+02]
 [-1.89272974e+01]
 [ 6.38766231e+00]
 [ 2.99181425e-01]
 [ 4.91180039e+02]
 [-2.07590556e+03]
 [ 6.04589329e+02]
 [-5.49887415e-01]
 [-2.10415158e+02]
 [ 4.16909313e+01]
 [-1.50370488e+00]
 [-1.28513004e+02]
 [ 1.52325225e+00]
 [-7.89080294e+01]]
```

（第一个是截距，其余的为 13 个权重）

Loss 值为：

```
21.763633492694172
```

使用解析法求得的系数求解测试集得到的结果为：

```
[[20.7909115  24.40900511 21.31874915 19.25335966 18.04289213 27.02204938
  36.51905509 28.64456151 13.62066846 17.11718498 21.64219289 18.59966739
   7.69062232 29.11794786 21.07391559  0.20918222 22.46254074 21.56078212
  23.79628503 21.0823101  11.54485976 19.5181533  24.48217801 20.96208081
  17.75648273 21.28637309 17.36485982 15.25560388 18.17409287  9.40810696
   8.56435044  5.03984438 25.04060066 10.89418196 23.68567072 19.66902106
  -6.07572459 20.64322002 34.76635085 27.13635449 25.38180351 35.15692771
  40.560521   39.93885877 21.3257883  34.244353   36.11522073 21.14804802
  12.91378363 38.11916007 13.28650706]]
```

分析结果发现：预测得到的结果和真实值相比大体相等，但又个别数据差值较大。

使用梯度下降法求得的系数情况如下

当 learning_rate=0.001 时，系数为：

```
[[  6.25506957]
 [ -7.80958676]
 [  9.79582316]
 [ -6.05627257]
 [ -0.99570275]
 [ -1.53647988]
 [  2.91800456]
 [ -0.23835584]
 [ -4.92689823]
 [  4.31276145]
 [ -1.91652672]
 [ -7.7585253 ]
 [  1.45351137]
 [-32.14622861]]
```

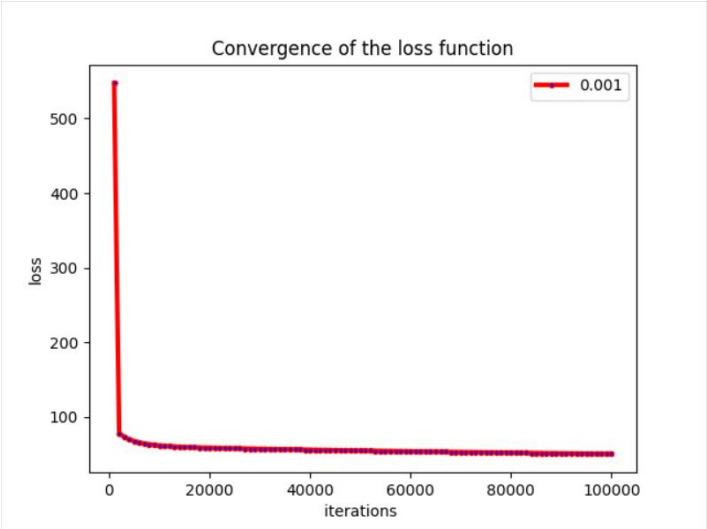
（第一个是截距，其余的为 13 个权重）

Loss 值为： 49.45973328287853

使用以上系数求解测试集得到的结果为：

```
[[21.84526564 24.17502308 22.84747017 17.16158687 25.14993648 25.18974958
  27.4960855  25.28566219 20.84882629 23.97184876 19.02730181 16.67118324
  14.15340737 25.57073755 25.10520245  9.55197337 26.30900845 22.61958417
  19.03377323 18.43521717 14.35557354 24.45424248 24.557681   17.2968643
  17.27228462 25.41641763 22.68857542 20.23035735 22.92823711 11.72961994
  11.32337599 13.58624657 25.25926248 14.16466281 19.1974833  16.56924571
   8.11366628 22.20605552 27.60437114 25.21742686 25.43474463 28.47289545
  25.63529036 20.17067326 19.39774135 28.44913346 31.2169751  22.28722285
  12.68766136 25.99872047 11.8574967 ]]
```

收敛过程的图像如下：



当 learning_rate=0.005 时，系数为：

```
[[  -8.2441174 ]
 [ -13.05922277]
 [   8.03460777]
 [  -5.06301766]
 [   8.29245024]
 [   5.46004885]
 [  24.03631932]
 [   6.52062124]
 [ -16.5352441 ]
 [  30.49405539]
 [  -2.086091  ]
 [ -28.83149877]
 [   1.21093079]
 [-102.7012996 ]]
```

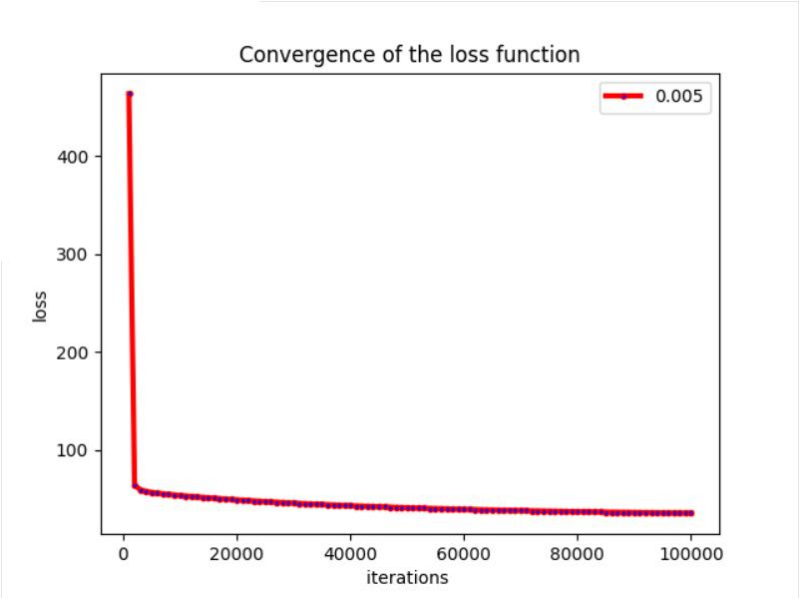
（第一个是截距，其余的为 13 个权重）

Loss 值为：`loss=34.21719129493289`

使用以上系数求解测试集得到的结果为：

```
[[20.79176338 24.51251381 21.29192968 17.87786267 24.65364171 27.00234267
  28.77407037 26.77674147 14.16758031 20.91699509 20.97657914 16.54011239
   8.38786088 28.60393982 23.80401368  2.57777368 27.3234599  22.84613996
  22.48905766 18.48210495 12.38703207 24.65304935 24.19503411 18.38254104
  17.13825638 25.13535791 19.40620875 18.16269461 20.63865211  9.69413597
   8.56257157  6.58978508 26.62243464  8.28708006 22.00192643 18.20586572
  -1.22226127 21.37408393 29.02242297 26.03433129 26.48671612 30.37888541
  30.73900088 27.19110229 21.21318011 30.65960015 32.32293891 20.96250531
  11.0251114  31.86539347  9.22199494]]
```

收敛过程的图像如下：



当 learning_rate=0.01 时，系数为：

```
[[ -16.88664098]
 [ -14.35080326]
 [   7.34391876]
 [   0.3148555 ]
 [  16.15131134]
 [   9.62525323]
 [  47.11876659]
 [   7.44141768]
 [ -41.35710375]
 [  46.4374473 ]
 [  -2.44094074]
 [ -62.99754551]
 [   1.21768969]
 [-127.60672005]]
```

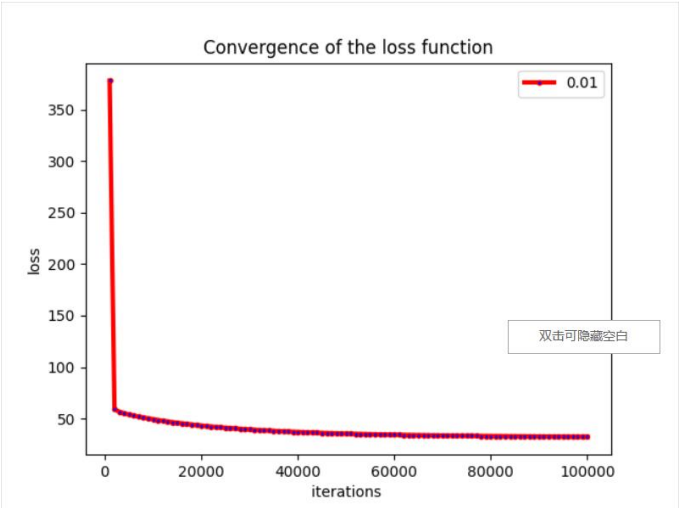
（第一个是截距，其余的为 13 个权重）

Loss 值为： loss=30.407381282871803

使用以上系数求解测试集得到的结果为：

```
[[20.42342318 24.63150764 20.64411756 18.17528711 23.49594636 27.88934499
  29.39973118 27.19880798 12.23054647 19.32160975 21.96215772 16.60075663
   6.42904199 29.8251643  23.0129299  -0.14275013 27.25927644 22.20534974
  23.82641271 19.28692999 11.66418261 24.08942722 23.81542757 18.84038551
  17.24586604 24.40223169 18.04704231 16.94466018 19.64948802  8.82521062
   7.41720389  4.26264182 26.8612533   6.48076051 23.17951594 18.79134368
  -4.65991255 20.82537134 29.94510311 26.25879481 26.73802245 32.0015451
  32.95049051 30.16108168 22.12584982 32.18909324 33.20013058 20.29650704
  10.33808473 34.19411975  8.28534734]]
```

收敛过程的图像如下：



当 learning_rate=0.05 时，系数为：

```
[[ -28.62557118]
 [ -17.05399933]
 [   7.14366836]
 [  -4.21601174]
 [  36.53176865]
 [  13.33426381]
 [ 124.61637282]
 [   3.39592284]
 [-120.04294125]
 [  52.61017971]
 [  -2.35986741]
 [-129.91853461]
 [   1.41152211]
 [-118.72569251]]
```

（第一个是截距，其余的为 13 个权重）

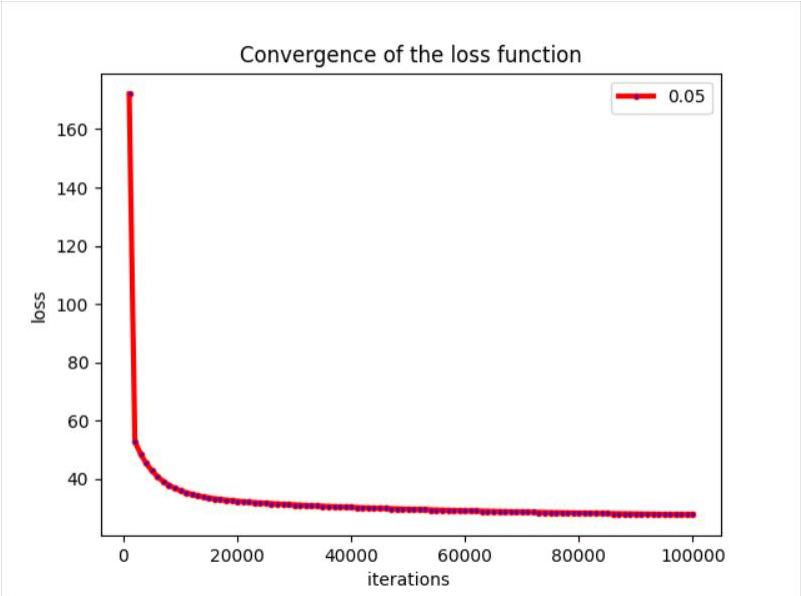
Loss 值为：

```
loss=26.57411915772434
```

使用以上系数求解测试集得到的结果为：

```
[[21.16836195 24.6205977  21.53377291 18.20949422 20.65417523 28.35163803
  29.85351354 27.62834602 12.00298084 18.1951936  22.2355241  17.09662012
   7.02754624 30.52986152 22.16256089 -0.32657833 25.46520425 21.14463523
  23.75933273 21.15950908 11.79540782 22.29894722 24.00171919 19.03441146
  17.74284377 22.43320886 17.45319305 15.53632568 19.26879757  8.67199573
   7.41247635  4.86736374 26.01028601  7.91013752 23.11334945 18.89506807
  -5.21851342 20.85810757 32.00226858 26.81733714 26.15682077 34.33823902
  34.47277747 31.56530869 22.14910827 33.83066959 34.92301369 19.75695391
  10.57924098 34.99708054  8.98000952]]
```

收敛过程的图像如下：



当 learning_rate=0.1 时，系统报错。

通过比较分析，发现使用解析法求得的系数拟合的 loss 最小，即拟合效果最好。分析梯度下降曲线发现：当 learning_rate=0.001 时初始误差函数下降得非常快，后突然变慢，其他情况下，误差函数的值下降速度较为均衡，没有初始时误差函数下降得非常快的情况，怀疑出现类蹦跳的情况。

结论与展望：

1. 提给条件的 learning_rate 是不合理的，我认为其应该随迭代次数的增加而降低；
2. 并不能确定波士顿房价与各个因素的关系就是线性的，所以可以尝试不同的误差函数来训练好的模型。