MultivariateLinearRegressionUsingNormalEquation

2022 年 3 月 18 日

```python
from csv import reader
import numpy as np

train_file=open("covid.train.csv")
rdr=reader(train_file)
a=[]
tar=[]
fst=True
for row in rdr:
    if fst:
        fst=False
    else:
        a.append(list(map(float,[1]+row[1:-1])))
        tar.append(float(row[-1]))
```

```python
x=np.array(a)
y=np.array(tar)
theta=np.linalg.pinv(x.T.dot(x)).dot(x.T).dot(y)
# print(theta)
print(theta.shape)
```

```
(94,)
```

```python
for i in range(y.shape[0]):
    print(y[i],theta.dot(x[i]))
```

```python
file=open("3.txt","w")
file.write(str(theta))
file.close()
```

```python
def J(theta):
    m=x.shape[0]
    res=0
    for i in range(m):
        tmp=theta.dot(x[i])-y[i]
        res+=tmp*tmp
    return res/(2*m)
J(theta)
```

```
0.41006181011388365
```

```python
test_file=open("covid.test.csv")
rdrt=reader(test_file)
b=[]
fst=True
for row in rdrt:
    if fst:
        fst=False
    else:
        b.append(list(map(float,[1]+row[1:])))
tx=np.array(b)
tx.shape
```

```
(893, 94)
```

```python
from csv import writer
outfile=open("result.csv","w")
wtr=writer(outfile,lineterminator='\n')
header=["id","tested_positive"]
wtr.writerow(header)
for i in range(tx.shape[0]):
    wtr.writerow([i,theta.dot(tx[i])])
outfile.close()
```