实验九（一）

线性回归示例代码

数据集下载：https://github.com/JerryKuan/data

# -\*- coding: UTF-8 -\*-

"""

此脚本用于实现线性回归模型

"""

​

​

import os

​

import numpy as np

import matplotlib.pyplot as plt

import pandas as pd

from sklearn import linear\_model

​

​

def read\_data(path):

"""

使用pandas读取数据

参数

----

path: String，数据的路径

返回

----

data: DataFrame，建模数据

"""

data = pd.read\_csv(path)

return data

​

​

def train\_model(x, y):

"""

利用训练数据，估计模型参数

参数

----

x: DataFrame，特征

y: DataFrame，标签

返回

----

model : LinearRegression, 训练好的线性模型

"""

# 创建一个线性回归模型

model = linear\_model.LinearRegression()

# 训练模型，估计模型参数

model.fit(x, y)

return model

​

​

def evaluate\_model(model, x, y):

"""

计算线性模型的均方差和决定系数

参数

----

model : LinearRegression, 训练完成的线性模型

x: DataFrame，特征

y: DataFrame，标签

返回

----

mse : np.float64，均方差

score : np.float64，决定系数

"""

# 均方差(The mean squared error)，均方差越小越好

mse = np.mean(

(model.predict(x) - y) \*\* 2)

# 决定系数(Coefficient of determination)，决定系数越接近1越好

score = model.score(x, y)

return mse, score

​

​

def visualize\_model(model, x, y):

"""

模型可视化

"""

# 创建一个图形框

fig = plt.figure(figsize=(6, 6), dpi=80)

# 在图形框里只画一幅图

ax = fig.add\_subplot(111)

ax.set\_xlabel('$x$')

ax.set\_ylabel('$y$')

# 画点图，用蓝色圆点表示原始数据

ax.scatter(x, y, color='b')

# 根据截距的正负，打印不同的标签

ax.plot(x, model.predict(x), color='r',

label=u'$y = %.3fx$ + %.3f' % (model.coef\_, model.intercept\_))

plt.legend(shadow=True)

# 展示上面所画的图片。图片将阻断程序的运行，直至所有的图片被关闭

# 在Python shell里面，可以设置参数"block=False"，使阻断失效。

plt.show()

​

​

def run\_model(data):

"""

线性回归模型建模步骤展示

​

参数

----

data : DataFrame，建模数据

"""

features = ["x"]

label = ["y"]

# 产生并训练模型

model = train\_model(data[features], data[label])

# 评价模型效果

mse, score = evaluate\_model(model, data[features], data[label])

print("MSE is %f" % mse)

print("R2 is %f" % score)

# 图形化模型结果

visualize\_model(model, data[features], data[label])

​

​

if \_\_name\_\_ == "\_\_main\_\_":

home\_path = os.path.dirname(os.path.abspath(\_\_file\_\_))

# Windows下的存储路径与Linux并不相同

if os.name == "nt":

data\_path = "%s\\simple\_example.csv" % home\_path

else:

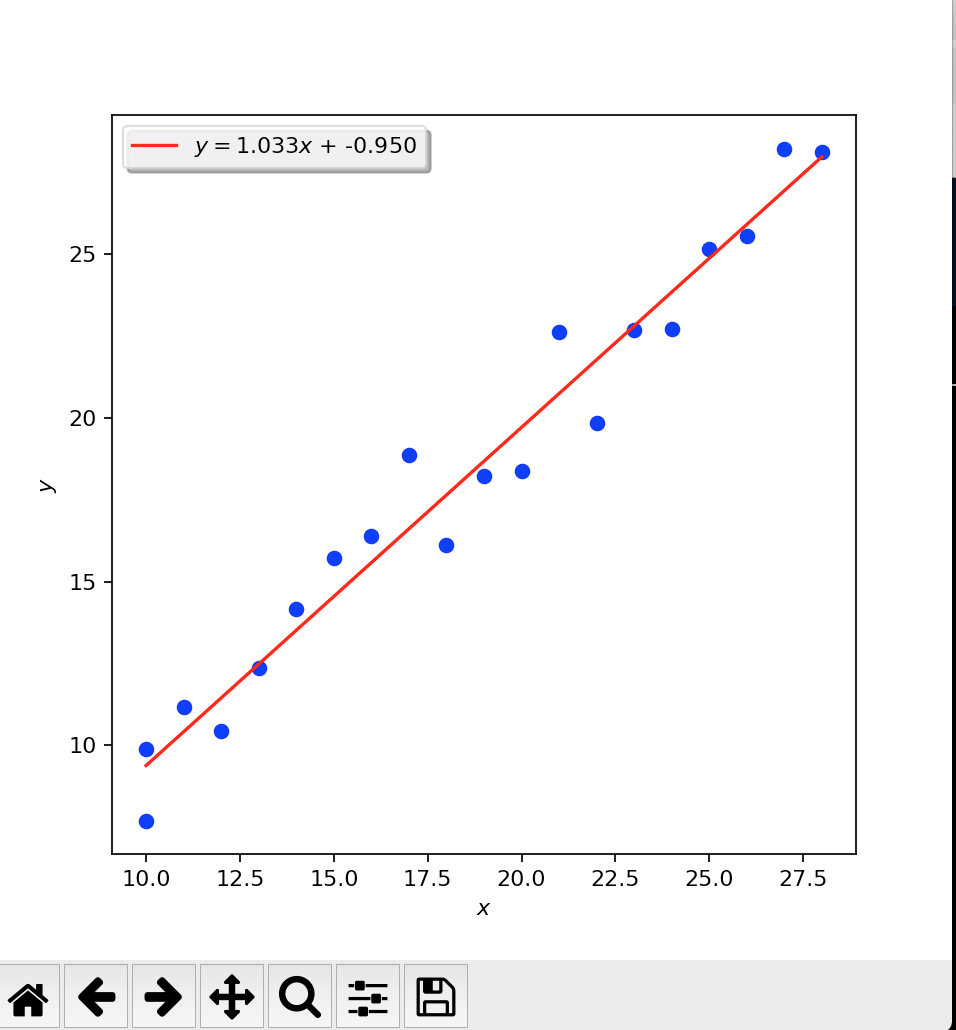
data\_path = "%s/simple\_example.csv" % home\_path

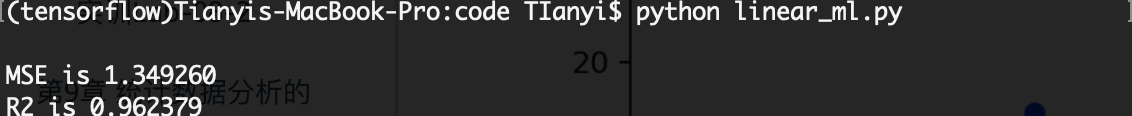
data = read\_data(data\_path)

run\_model(data)



实验结果

​​

​​

​