

Python作业第二周

项目发起人: [江柳]

项目发起时间: [2017.10.29]

- 1.课程阅读
- 2.作业布置
 - 2.1 数据载入
 - 2.2 代价计算
 - 2.3 梯度下降算法
 - 2.4 预测
- 3.Deadline
- 4.提交方法

1.课程阅读

machine learning

本次作业目标:

假设你是海底捞的CEO。不同城市的分店具有不同的客流量，与此对应的盈利也不一样。现在你手上有本公司97家连锁超市的客流信息和盈利信息[在文本dataset_2_1.txt中]。现在公司计划新家一家连锁店，有两个不同客流量的城市可以选择，一个客流量为3.5万人，另一个城市客流量为7万人，试预测如果在两个城市开分店，分别能盈利多少元？

涉及内容:

- Linear Regression模型
- 损失函数的计算
- 函数梯度的计算

注:

- 本次作业全部代码在 `ex2.py` 文件中完成

2.作业布置

2.1 数据载入

文本 `dataset_2_1.txt` 中有两列数据，第一列客流量，单位万人，第二列表示对应的盈利，单位万元。

尝试补全`load(filename)`函数代码，使该函数能读取 `dataset_2_1.txt`，并返回列表 `x, y`。

完成后直接运行 `ex2.py` 文件，输出结果如下:

```
Loading Data ...
```

```
m = 97
```

```
X[:10] = [6.1101, 5.5277, 8.5186, 7.0032, 5.8598, 8.3829, 7.4
```

```
y[:10] = [17.592, 9.1302, 13.662, 11.854, 6.8233, 11.886, 4.3
```

2.2 代价计算

我们假设数据满足模型:

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

最接近现实情况的模型，我们需要衡量什么叫最接近现实情况，在这里我们定义一个损失函数用来衡量模型与现实情况的近似程度。

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^i) - y^i)^2$$

其中 m 是样本的个数， x^i 和 y^i 是第 i 个样本， $J(\theta)$ 表示损失大小。

尝试补全 `computeCost(X, y, theta0, theta1)` 函数代码，使该函数能够根据参数返回损失数值。

完成后直接运行 `ex2.py` 文件，输出结果如下：

```
Testing CostFuntion ...  
With theta0 = 0 and theta1 = 0  
Cost computed = 32.072734  
Expected cost value (approx) 32.07  
with theta0 = -1 and theta1 = 2  
Cost computed = 54.242455  
Expected cost value (approx) 54.24
```

2.3 梯度下降算法

现在我们已经有了损失函数了，求最接近现实情况的模型，就是求使 $J(\theta)$ 最小的 θ_0 和 θ_1 。这里已经变成一个参数优化问题了，解参数优化问题的方法很多，可以自行了解，这里选用随机梯度下降算法 (Gradient Descent)。

- 随机梯度下降算法
 - 随机选择参数 θ_0 和 θ_1
 - 沿梯度反方向下降，重复此过程，知道损失函数收敛。

即：

$$\theta_0 = \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^i) - y^i)$$

$$\theta_1 = \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^i) - y^i) x^i$$

注:

- θ_0 和 θ_1 要同时变化。

尝试补全 `gradientDescent(X, y, theta0, theta1, alpha, num_iterations)` 函数代码, 使该函数返回模型对应的 `theta0` 和 `theta1` 。

完成后直接运行 `ex2.py` 文件, 输出结果如下:

```
Running Gradient Descent ...
Theta found by gradient descent:
theta0 = -3.630291, theta1 = 1.166362
Expected theta values (approx)
theta0 = -3.6303, theta1 = 1.1664
```

2.4 预测

该部分不需要补充代码, 直接运行 `ex2.py` 文件即可:

```
For population = 35000, we predict a profit of 4519.77
For population = 70000, we predict a profit of 45342.45
```

3. Deadline

- 2017.11.8上午12:00截止
- 有不懂得可以随时Google或者找我问

4. 提交方法

- 邮件发送到: `[root@oopy.org]`
- 邮件标题: (姓名全拼) lecture2

- 作业以邮件附件形式发送