

吴恩达Coursera深度学习课程 DeepLearning.ai 提炼笔记 (1-2) -- 神经网络基础

原创 2017年09月20日 21:14:18 标签: deep-learning / 吴恩达 / Coursera / 神经网络和深度学习

7967



Koala_Tree

原创 171 粉丝 569 喜欢 184



等级: 博客 访问量: 17
积分: 3457 排名: 1万+

作者: 大树先生

博客: http://blog.csdn.net/koala_tree

GitHub: <https://github.com/KoalaTree>

2017年09月20日

以下为在Coursera上吴恩达老师的DeepLearning.ai课程项目中, 第一部分《神经网络和深度学习》第二周课程部分关键点的笔记。笔记并不包含全部小视频课程的记录, 如需学习笔记中舍弃的内容请至Coursera 或者 网易云课堂。同时在阅读以下笔记之前, 强烈建议先学习吴恩达老师的视频课程。

同时我在知乎上开设了关于机器学习深度学习的专栏收录下面的笔记, 方便在移动端的学习。欢迎关注我的知乎: 大树先生。一起学习一起进步呀! ^_^

神经网络和深度学习—神经网络基础

1. 二分类问题

对于二分类问题, 大牛给出了一个小小的Notation。

- 样本: (x, y) , 训练样本包含 m 个;
- 其中 $x \in R^{n_x}$, 表示样本 x 包含 n_x 个特征;
- $y \in 0, 1$, 目标值属于0、1分类;
- 训练数据: $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$

输入神经网络时样本数据的形状:

$$X.shape = (n_x, m)$$

目标数据的形状:

$$Y = [y_{(1)}, y_{(2)}, \dots, y_{(m)}]$$

$$Y.shape = (1, m)$$

2. logistic Regression

逻辑回归中, 预测值:

$$\hat{h} = P(y = 1|x)$$

其表示为1的概率, 取值范围在 $[0, 1]$ 之间。



博主最新文章

完结撒花! 吴恩达DeepLearning度学习》课程笔记目录总集

LeetCode-54, 59: Spiral Matrix (输出矩阵)

LeetCode-62, 63: Unique Paths (中独一无二的路径)

吴恩达 深度学习 编程作业 (5-3) - Trigger word detection

吴恩达 深度学习 编程作业 (5-3) - Neural Machine Translation

文章分类

Deep Learning

TensorFlow

算法

吴恩达 深度学习 课程笔记

吴恩达 深度学习 编程作业

LeetCode刷题

展开

博主专栏



剑指Offer 笔试面试题
74篇 10627



16篇

吴恩达《深度学习》笔记
44363

LeetCode刷题
29683

引入Sigmoid函数，预测值：

$$\hat{y} = \text{Sigmoid}(w^T x + b) = \sigma(w^T x + b)$$

其中

$$\text{Sigmoid}(z) = \frac{1}{1 + e^{-z}}$$

注意：函数的一阶导数可以用其自身表示，

$$\sigma'(z) = \sigma(z)(1 - \sigma(z))$$

这里可以解释梯度消失的问题，当 $z = 0$ 时，导数最大，但是导数最大为 $\sigma'(0) = \sigma(0)(1 - \sigma(0)) = 0.5(1 - 0.5) = 0.25$ ，这里导数仅为原函数值的0.25倍。

参数梯度下降公式的不断更新， $\sigma'(z)$ 会变得越来越小，每次迭代参数更新的步伐越来越小，最终接近于0，产生梯度消失的现象。

3. logistic回归 损失函数

Loss function

一般经验来说，使用平方错误（squared error）来衡量Loss Function：

$$L(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$$

但是，对于logistic regression 来说，一般不适用平方错误来作为Loss Function，这是因为上面的平方错误损失函数一般是非凸函数（non-convex），其在使用低度下降算法的时候，容易得到局部最优解，而不是全局最优解。因此要选择凸函数。

逻辑回归的Loss Function：

$$L(\hat{y}, y) = -(y \log \hat{y} + (1 - y) \log(1 - \hat{y}))$$

- 当 $y = 1$ 时， $L(\hat{y}, y) = -\log \hat{y}$ 。如果 \hat{y} 越接近1， $L(\hat{y}, y) \approx 0$ ，表示预测效果越好；如果 \hat{y} 越接近0， $L(\hat{y}, y) \approx +\infty$ ，表示预测效果越差；
- 当 $y = 0$ 时， $L(\hat{y}, y) = -\log(1 - \hat{y})$ 。如果 \hat{y} 越接近0， $L(\hat{y}, y) \approx 0$ ，表示预测效果越好；如果 \hat{y} 越接近1， $L(\hat{y}, y) \approx +\infty$ ，表示预测效果越差；
- 我们的目标是最小化样本点的损失Loss Function，损失函数是针对单个样本点的。

Cost function

全部训练数据集的Loss function总和的平均值即为训练集的代价函数（Cost function）。

$$J(w, b) = \frac{1}{m} \sum_{i=1}^m L(\hat{y}^{(i)}, y^{(i)}) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)})]$$

- Cost function是待求系数w和b的函数；
- 我们的目标就是迭代计算出最佳的w和b的值，最小化Cost function，让其尽可能地接近于0。

4. 梯度下降

用梯度下降法（Gradient Descent）算法来最小化Cost function，以计算出合适的w和b的值。

每次迭代更新的修正表达式：

$$w := w - \alpha \frac{\partial J(w, b)}{\partial w}$$

$$b := b - \alpha \frac{\partial J(w, b)}{\partial b}$$

在程序代码中，我们通常使用dw来表示 $\frac{\partial J(w, b)}{\partial w}$ ，用db来表示 $\frac{\partial J(w, b)}{\partial b}$ 。

5. 逻辑回归中的梯度下降法

对单个样本而言，逻辑回归Loss function表达式：



文章存档

2018年4月

2018年3月

2018年2月

2018年1月

2017年12月

2017年11月

展开

博主热门文章

吴恩达 深度学习 编程作业（1-2）on Basics with Numpy & Logistic Regression

吴恩达 深度学习 编程作业（1-4）Building your DNN & DNN for Image Classification

吴恩达 Coursera 深度学习课程 De ning.ai 提炼笔记（1-2）-- 神经网络

吴恩达 深度学习 编程作业（1-3）r data classification with one hidden layer

卷积神经网络循环神经网络结合-CNN+RNN

吴恩达 Coursera 深度学习课程 De ning.ai 提炼笔记（2-1）-- 深度学习

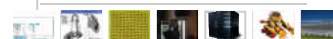
吴恩达 Coursera 深度学习课程 De ning.ai 提炼笔记（1-3）-- 浅层神经网络

吴恩达 Coursera 深度学习课程 De ning.ai 提炼笔记（4-1）-- 卷积神经网络

吴恩达 深度学习 编程作业（2-2）mization Methods

Kaggle_Titanic生存预测 -- 详细笔记梳理

废机油提炼柴油



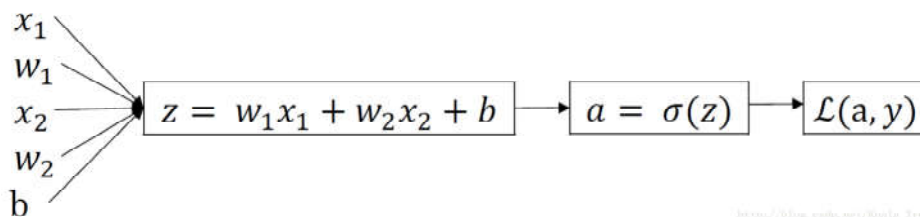
联系我们



请扫描二维码联系
webmaster@
400-660-0106
QQ客服

关于 招聘 广告服务

反向传播过程:

$$z = w^T x + b$$
$$\hat{y} = a = \sigma(z)$$
$$L(a, y) = -(y \log(a) + (1 - y) \log(1 - a))$$


前面过程的da、dz求导:

$$da = \frac{\partial L}{\partial a} = -\frac{y}{a} + \frac{1-y}{1-a}$$
$$dz = \frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} = \left(-\frac{y}{a} + \frac{1-y}{1-a}\right) \cdot a(1-a) = a - y$$

再对 w_1 、 w_2 和b进行求导:

$$dw_1 = \frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial w_1} = x_1 \cdot dz = x_1(a - y)$$
$$db = \frac{\partial L}{\partial b} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial b} = 1 \cdot dz = a - y$$

梯度下降法:

$$w_1 := w_1 - \alpha dw_1$$

$$w_2 := w_2 - \alpha dw_2$$

$$b := b - \alpha db$$

6. m个样本的梯度下降

对m个样本来说，其Cost function表达式如下:

$$z^{(i)} = w^T x^{(i)} + b$$
$$\hat{y}^{(i)} = a^{(i)} = \sigma(z^{(i)})$$
$$J(w, b) = \frac{1}{m} \sum_{i=1}^m L(\hat{y}^{(i)}, y^{(i)}) = -\frac{1}{m} \sum_{i=1}^m \left[y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)}) \right]$$

Cost function 关于w和b的偏导数可以写成所有样本点偏导数和的平均形式:

$$dw_1 = \frac{1}{m} \sum_{i=1}^m x_1^{(i)} (a^{(i)} - y^{(i)})$$
$$db = \frac{1}{m} \sum_{i=1}^m (a^{(i)} - y^{(i)})$$

7. 向量化 (Vectorization)

在深度学习的算法中，我们通常拥有大量的数据，在程序的编写过程中，应该尽最大可能的少使用loop循环语句，利用python可以实现矩阵运算，进而来提高程序的运行速度，避免for循环的使用。

逻辑回归向量化

- 输入矩阵X: (n_x, m)
- 权重矩阵w: $(n_x, 1)$
- 偏置b: 为一个常数

- 输出矩阵Y：
(1, m)

所有m个样本的线性输出Z可以用矩阵表示：

$$Z = w^T X + b$$

python代码：

```
1 Z = np.dot(w.T,X) + b
2 A = sigmoid(Z)
```

逻辑回归梯度下降输出向量化

- dZ对于m个样本，维度为 (1, m) ，表示为：

$$dZ = A - Y$$

- db可以表示为：

$$db = \frac{1}{m} \sum_{i=1}^m dz^{(i)}$$

python代码：

```
1 db = 1/m*np.sum(dZ)
```

- dw可表示为：

$$dw = \frac{1}{m} X \cdot dZ^T$$

python代码：

```
1 dw = 1/m*np.dot(X,dZ.T)
```

单次迭代梯度下降算法流程

```
1 Z = np.dot(w.T,X) + b
2 A = sigmoid(Z)
3 dZ = A-Y
4 dw = 1/m*np.dot(X,dZ.T)
5 db = 1/m*np.sum(dZ)
6
7 w = w - alpha*dw
8 b = b - alpha*db
```

8. python的notation

- 虽然在Python有广播的机制，但是在Python程序中，为了保证矩阵运算的正确性，可以使用reshape()函数来对矩阵设定所需要进行计算的维度，这是个好的习惯；
- 如果用下列语句来定义一个向量，则这条语句生成的a的维度为 (5,)，既不是行向量也不是列向量，称为秩 (rank) 为1的array，如果对a进行转置，则会得到a本身，这在计算中会给我们带来一些问题。

```
1 a = np.random.randn(5)
```

- 如果需要定义 (5, 1) 或者 (1, 5) 向量，要使用下面标准的语句：

```
1 a = np.random.randn(5,1)
2 b = np.random.randn(1,5)
```

- 可以使用assert语句对向量或数组的维度进行判断。assert会对内嵌语句进行判断，即判断a的维度是不是 (5, 1)，如果不是，则程序在此处停止。使用assert语句也是一种很好的习惯，能够帮助我们及时检查、发现语句是否正确。

```
1 assert(a.shape == (5,1))
```

- 可以使用reshape函数对数组设定所需的维度

```
1 a.reshape((5,1))
```

8. logistic regression代价函数的解释

Cost function的由来:

预测输出 \hat{y} 的表达式:

$$\hat{y} = \sigma(w^T x + b)$$

其中,

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

。

\hat{y} 可以看作预测输出为正类 (+1) 的概率:

$$\hat{y} = P(y = 1|x)$$

当 $y = 1$ 时, $P(y|x) = \hat{y}$; 当 $y = 0$ 时, $P(y|x) = 1 - \hat{y}$ 。

将两种情况整合到一个式子中, 可得:

$$P(y|x) = \hat{y}^y (1 - \hat{y})^{(1-y)}$$

对上式进行log处理 (这里是因为log函数是单调函数, 不会改变原函数的单调性):

$$\log P(y|x) = \log \left[\hat{y}^y (1 - \hat{y})^{(1-y)} \right] = y \log \hat{y} + (1 - y) \log(1 - \hat{y})$$

概率 $P(y|x)$ 越大越好, 即判断正确的概率越大越好。这里对上式加上负号, 则转化成了单个样本的Loss function, 我们期望其值越小越好:

$$L(\hat{y}, y) = -(y \log \hat{y} + (1 - y) \log(1 - \hat{y}))$$

对于m个训练样本来说, 假设样本之间是独立同分布的, 我们总是希望训练样本判断正确的概率越大越好, 则有:

$$\max \prod_{i=1}^m P(y^{(i)}|x^{(i)})$$

同样引入log函数, 加负号, 则可以得到Cost function:

$$J(w, b) = \frac{1}{m} \sum_{i=1}^m L(\hat{y}^{(i)}, y^{(i)}) = -\frac{1}{m} \sum_{i=1}^m \left[y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)}) \right]$$

本周 (Week2) 的课后编程题请参见:

[吴恩达Coursera深度学习课程 DeepLearning.ai 编程作业 \(1-2\)](#)

版权声明: 本文为博主原创文章, 未经允许不得转载。 https://blog.csdn.net/Koala_Tree/article/details/78045596

本文已收录于以下专栏: [吴恩达《深度学习》课程笔记](#)



[查看更多>>](#)

看 Python 如何诠释 “薪” 时代

Python全栈开发包含Python爬虫、前端、网站后台、Python机器学习与数据挖掘等, 从0基础小白到Python 企业级web开发达人、自动化运维开发能手的进阶, 课程真实企业项目实战演练, 全面系统学习python编程语言, 从容应对企业中各式各样的.....

29224

[查看更多>>](#)



目前您尚未登录, 请 [登录](#) 或 [注册](#) 后进行评论



Willen_ 2018-04-06 16:01 #10楼

[回复](#)

真的厉害,感谢楼主!1



zhanglianpin 2018-03-12 11:10 #9楼

[回复](#)

非常好的资料，谢谢楼主整理分享。谢谢，谢谢，谢谢。重要的事情说3遍。



zcsdn_2017 2018-02-23 13:01 #8楼

[回复](#)

非常感谢，有笔记，还有练习与答案。

[查看 13 条热评](#)▼

Coursera吴恩达《神经网络与深度学习》课程笔记（2）-- 神经网络基础之逻辑回归

上节课我们主要对深度学习（Deep Learning）的概念做了简要的概述。我们先从房价预测的例子出发，建立了标准的神经网络（Neural Network）模型结构。然后从监督式学习入手，介绍了Sta...



red_stone1 2017年09月07日 18:06 5387

Coursera吴恩达《神经网络与深度学习》课程笔记（1）-- 深度学习概述

我的CSDN博客地址：红色石头的专栏 我的知乎主页：红色石头 我的微博：RedstoneWill的微博 我的GitHub：RedstoneWill的GitHub 我的微信...



red_stone1 2017年09月03日 10:31 9429

程序员提升自己100种方法，我推荐这种！

给自己定下学习目标，并坚持下去，结果可能就不一样了！



吴恩达【深度学习工程师】学习笔记（二）



zchang81 2017年09月14日 11:19 2896

1、二分类问题； 2、逻辑回归及其对应的代价函数形式； 3、用计算图描述神经网络的正向、反向传播过程； 4、在逻辑回归中使用梯度下降算法。...

深度学习DeepLearning.ai系列课程学习总结：1. 深度学习简介

本节内容对应吴恩达deeplearning.ai课程中，第一部分第一周的内容。学习过程中的一些笔记与心得与大家分享！原始课程请访问<https://www.deeplearning.ai/>来了解...



laodengbaiwe0838 2017年08月20日 21:49 5711

吴恩达机器学习笔记_第一周



hunterlew 2016年04月02日 16:51 5018

毕业论文方向可能和神经网络挂钩，神经网络也是机器学习的一部分。从这周开始决定跟着Andrew Ng公开课系统地学习机器学习，4/4日开课，第一周为试听，今天先看了。以后每周更新一下。的确感觉讲的不错，...

提炼黄金设备

提炼黄金需要用哪些设备

百度广告



Coursera深度学习课程笔记 deep learning

2017年09月08日 16:53 5.12MB

[下载](#)



吴恩达Coursera深度学习课程 DeepLearning.ai 提炼笔记（4-3）-- 目标检测

卷积神经网络 --- 目标检测

吴恩达DeepLearning.ai笔记 (5-1) -- 循环序列模型

吴恩达DeepLearning.ai笔记 (5-1) - 循环序列模型 1.一些序列数据例子 2.数学符号 x x^1 x^2 ...

 sinat_28002567 2018年03月01日 16:50  236

deep learning.ai 序列模型笔记

 owenfy 2018年03月11日 21:20  18

1、应用：2、notation人名识别X: Harry Potter and Hermione Granger invented a new spell. x^1 x^2 ...

吴恩达 深度学习 编程作业 (1-4) - Building your DNN & DNN for Image Classifi...

Part 1: Building your Deep Neural Network: Step by Step Part 2: Deep Neural Network for Image Classi...

 Koala_Tree 2017年09月26日 10:18  10417



技术外文文献看不懂？教你一个公式秒懂英语
不背单词和语法，一个公式学好英语

Coursera 吴恩达DeepLearning.AI 第五课 sequence model 序列模型 第一周Buildi...

本周的习题有点多，主要是python不熟悉，然后时间不够，提醒说马上过期才开始看的视频，optional部分没有写完Building your Recurrent Neural Network - S...

 forqzy 2018年02月06日 07:34  1670

Coursera 吴恩达DeepLearning.AI 第五课 sequence model 序列模型 第二周 Opera...

只做了计分部分Operations on word vectorsWelcome to your first assignment of this week!Because word embeddin...

 forqzy 2018年02月12日 10:39  999

吴恩达 深度学习 编程作业 (5-2) Part 2 - Emojify!

Emojify: many-to-one

 Koala_Tree 2018年03月06日 11:30  690

Coursera 吴恩达DeepLearning.AI 第五课 sequence model 序列模型 第一周 Impro...

本周的习题有些地方有点坑，主要是题目讲述的不够清楚，以下答案供参考Improvise a Jazz Solo with an LSTM NetworkWelcome to your final pro...

 forqzy 2018年02月06日 07:22  1262

Coursera 吴恩达DeepLearning.AI 第五课 sequence model 序列模型 第二周 Emofify

这个Emojify里最坑的一个就是，avg初始化的时候一定要是 (50,)，如果你用(word_to_vec_map["a"]).shape 就死活过不了。Emojify!Wel...

 forqzy 2018年02月12日 10:33  807

脸部高原红如何去除

高原红怎么去除

百度广告



Coursera 吴恩达DeepLearning.AI 第五课 sequence model 序列模型 第一周 随堂测验

 justry24 2018年02月05日 21:04  1555

吴恩达 深度学习 编程作业 (2-2) - Optimization Methods

Optimization Methods

 Koala_Tree 2017年10月12日 16:26  4337

笔记 | 吴恩达Coursera Deep Learning学习笔记

向AI转型的程序员都关注了这个号👉👉👉作者: Lisa Song微软总部云智能高级数据科学家, 现居西雅图。具有多年机器学习和深度学习的应用经验, 熟悉各种业务场景下机器学习和人工智能产品的需求分析、架构设...

 dQCFKyQDXYm3F8rB0 2017年09月27日 00:00  1676

吴恩达Coursera深度学习课程 DeepLearning.ai 提炼笔记 (2-1) -- 深度学习的实践...

改善深层神经网络: 超参数调试、正则化以及优化 ---- 深度学习的实践方面

 Koala_Tree 2017年09月28日 17:00  4879

超级干货 | 优美的课程笔记, 吴恩达点赞的深度学习课程信息图

吴恩达在推特上展示了一份由 TessFerrandez 完成的深度学习专项课程信息图, 这套信息图优美地记录了深度学习课程的知识与亮点。因此它不仅仅适合初学者了解深度学习, 还适合机器学习从业者和研究者复...

 R1uNW1W 2018年03月08日 00:00  202
