

E14 BP Algorithm (C++/Python)

20214966 Yangkai Lin 20214810 Suixin Ou

December 12, 2020

Contents

1	Horse Colic Data Set	2
2	Reference Materials	2
3	Tasks	2
4	Codes and Results	3

1 Horse Colic Data Set

The description of the horse colic data set (<http://archive.ics.uci.edu/ml/datasets/Horse+Colic>) is as follows:

Data Set Characteristics:	Multivariate	Number of Instances:	368	Area:	Life
Attribute Characteristics:	Categorical, Integer, Real	Number of Attributes:	27	Date Donated	1989-08-06
Associated Tasks:	Classification	Missing Values?	Yes	Number of Web Hits:	108569

We aim at trying to predict if a horse with colic will live or die.

Note that we should deal with missing values in the data! Here are some options:

- Use the feature's mean value from all the available data.
- Fill in the unknown with a special value like -1.
- Ignore the instance.
- Use a mean value from similar items.
- Use another machine learning algorithm to predict the value.

2 Reference Materials

1. Stanford: **CS231n: Convolutional Neural Networks for Visual Recognition** by Fei-Fei Li, etc.
 - Course website: <http://cs231n.stanford.edu/2017/syllabus.html>
 - Video website: https://www.bilibili.com/video/av17204303/?p=9&tdsourcetag=s_pctim_aiomsg
2. **Machine Learning** by Hung-yi Lee
 - Course website: <http://speech.ee.ntu.edu.tw/~tlkagk/index.html>
 - Video website: <https://www.bilibili.com/video/av9770302/from=search>
3. A Simple neural network code template

3 Tasks

- Given the training set `horse-colic.data` and the testing set `horse-colic.test`, implement the BP algorithm and establish a neural network to predict if horses with colic will live or die. In addition, you should calculate the accuracy rate.
- Please submit a file named `E14.YourNumber.pdf` and send it to `ai_2020@foxmail.com`

- Draw the training loss and accuracy curves
- (optional) You can try different structure of neural network and compare their accuracy and the time they cost.

4 Codes and Results

- 代码见BP.py，我实现了单个隐藏层的训练batch=1的BP神经网络。
- 结果及分析见BP.html或BP.ipynb，我调节了学习率和隐藏层大小，观察不同的结果。