



Deep Learning (Homework 1)

Due date : 04/12/2019

- Any tools for **automatic differentiation** are **forbidden** in this homework, such as **Tensorflow**, **Pytorch**, **Keras**, **MXNet**, et cetera. You should implement the backpropagation **by yourself**.
- **Submitting Homework** – Please zip each of your **source code** and **report** into a single compress file and name the file using this format : **HW1_StudentID_StudentName.zip** (rar, 7z, tar.gz, ...etc are all not acceptable)

Deep Neural Network for Classification

In this exercise, please implement a **Deep Neural Network (DNN)** model to predict if one can survive in the Titanic tragedy. This dataset is collected from a Kaggle competition and cleaned by TAs. The features include: **sex**, **age** of a person; how many **parents** or **children**, and how many **siblings** or **spouses** are on board related to the person; the **ticket class** and the **passenger fare** are also included. Please create a DNN model **inputs** these features and **outputs** the **survival** of passengers. The details of this dataset can be referred to: <https://www.kaggle.com/c/titanic/data>. The cleaned data is provided in **titanic.csv**, please take the first 800 rows as training data and the last 91 rows as test data. The first column, the survival, should be treated as label, while the other six columns are features. In the sex column, 1 indicates male while 0 indicates female. In the age column, the empty entries are filled with zeros.

1. Please construct a DNN for binary classification. For N samples and K categories, the cross entropy objective function is expressed by

$$E(\mathbf{w}) = - \sum_{n=1}^N \sum_{k=1}^K t_{nk} \ln y_k(\mathbf{x}_n, \mathbf{w}).$$

Please minimize the objective function $E(\mathbf{w})$ by **error backpropagation** algorithm using the **Stochastic Gradient Descent** (SGD) where the parameter updating in each mini-batch is yielded by

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} - \eta \nabla E(\mathbf{w}^{(\tau)}).$$

You should decide the following variables: number of hidden layers, number of hidden units for each layer, learning rate, number of iterations and mini-batch size. Please show (a) **learning curve**, (b) **training error rate** and (c) **test error rate** in the report. The network architecture should be designed by yourself. Please describe the architecture of your model and explain why you design or choose such an architecture.

2. Please construct a DNN with a specified architecture. The number of units in **layers** should be arranged as [6 3 3 2], corresponding to those of input layer, first and second hidden layers, and output layer. This network has the same inputs and outputs as provided.
3. Look into the features. Please perform normalization for the **Fare** feature and compare the error rate with the original network. For normalization process, you may refer to https://en.wikipedia.org/wiki/Standard_score. Are there any other features that need to be normalized? Please answer in the report.

4. Please identify which **feature** affects the prediction performance the **most**. Describe the process how you solve this problem in the report.
5. When the feature is **categorical** with more than three categories, we often transform the categories into one-hot vectors. Should we perform the same process to the feature of **ticket class**? Please show your explanations and illustrations in the report.
6. Please artificially design two **new** samples that one can **survive** and the other can not. For example, a person with the following feature: pclass=3, sex=1, age=25, sibsp=2, parch=2, fare=10. Input this new sample to the model and see the result is survived or not. Explain how you choose these samples to achieve correction classification.

