Machine Learning-Homework 1

Shih-Min, Lin
Department of Applied Mathematics
National Chung Hsing University
Taichung, Taiwan

I. Formula Derivation

這次作業用了3-layer multilayer perceptron. 以下是Forward Propagation和Backpropagation的推導

A. *Forward Propagation*

$a_i^{(in)}$為第i個units in the input layers, $z_1^{(h)}$為第1個output data, $w_{i,1}^{(h)}$為第i個units in the input layers產生出第1個輸出的權重。

$$z_1^{(h)} = [w_{1,1}^{(h)} w_{2,1}^{(h)} \cdots w_{n,1}^{(h)}]^T [a_1^{(in)} a_2^{(in)} \cdots a_n^{(in)}] \quad (1)$$

在這裡我們把$\vec{z}$ 定義為$[z_1^{(h)} z_2^{(h)} \cdots z_n^{(h)}]^T$所以下式為$z$的定義

$$\begin{aligned} \vec{z}^{(h)} &= [z_1^{(h)} z_2^{(h)} \cdots z_n^{(h)}]^T \\ &= [w_1^{(h)} w_2^{(h)} \cdots w_n^{(h)}]^T \vec{a}^{(in)} \\ &= (W^h)^T \vec{a}^{(in)} \end{aligned} \quad (2)$$

ϕ定義為activation function,$\phi = \frac{1}{1+e^{-x}}$, 則$\vec{a}^{(h)} = \phi(\vec{z}^{(h)})$,接著我們把上面的式子寫成下列形式

$$Z^h = (W^{(h)})^T A^{(in)} \quad (3)$$

$$A^h = \phi(Z^{(h)}) \quad (4)$$

接著output layer可以寫成

$$Z^{out} = (W^{(out)})^T A^{(h)} \quad (5)$$

$$A^{out} = \phi(Z^{(out)}) \quad (6)$$

B. *Backpropagation*

$Lossfunction = \sum_i \left\|\hat{y}^i - y^i\right\|^2$,接著我們把loss function展開

$$\begin{aligned} Lossfunction &= \sum_i \left\|\hat{y}^i - y^i\right\|^2 \\ &= \sum_i \left\|\vec{a}^i - \vec{y}^i\right\|^2 \\ &= Tr(A^{(out)} - Y)^T (A^{(out)} - Y) \end{aligned} \quad (7)$$

把loss function對W_j做偏微分

$$\begin{aligned} &\tfrac{\partial}{\partial w_j} Tr(A^{(out)} - Y)^T (A^{(out)} - Y) \\ &= \frac{\partial Tr(A^{(out)} - Y)^T (A^{(out)} - Y)}{\partial A^{(out)} - Y} \frac{\partial A^{(out)} - Y}{\partial w_j} \\ &= 2(A^{(out)} - Y) \frac{\partial (A^{(out)} - Y)}{\partial A^{(out)}} \frac{\partial A^{(out)}}{\partial w_j} \\ &= 2(A^{(out)} - Y) I \frac{\partial ((W^{(out)})^T A^h)}{\partial W_j^{(out)}} \\ &= 2A_j^h \delta^{out} \quad , \delta^{out} = A^{out} - Y \end{aligned} \quad (8)$$

接著再把第(8)式對$W_{j,k}^h$做偏微分

$$\begin{aligned} &\tfrac{\partial Tr(A^{(out)} - Y)^T (A^{(out)} - Y)}{\partial w_{j,k}^h} \\ &= \frac{\partial Tr(A^{(out)} - Y)^T (A^{(out)} - Y)}{\partial (A^{(out)} - Y)} \frac{\partial (A^{(out)} - Y)}{\partial W_{i,j}^h} \\ &= 2(A - Y) \frac{\partial (A^{(out)} - Y)}{\partial A^{(out)}} \frac{\partial (A^{(out)})}{\partial W_{i,j}^h} \\ &= 2(A^{(out)} - Y) I \frac{\partial (W^{(out)})^T A^h}{\partial A^h} \frac{\partial A^h}{\partial W_{j,k}^h} \\ &= 2(A^{(out)} - Y) I \frac{\partial (W^{(out)})^T A^h}{\partial A^h} \frac{\partial \phi(Z^h)}{\partial Z^h} \frac{\partial Z^h}{\partial W_{j,k}^h} \\ &= 2(A^{(out)} - Y) I \frac{\partial (W^{(out)})^T A^h}{\partial A^h} [\phi(Z^h) \odot (C - \phi(Z^h))] \frac{\partial ((W^h)^T) A^{in}}{\partial W_{j,k}^h} \\ &= 2A_k^{in} \delta_j^h \qquad , \delta_j^h = W^{out} \delta(out) \odot \frac{\partial \phi(Z^h)}{\partial Z^h} \end{aligned} \quad (9)$$

II. Introduce house data and Preprocess

(1)Introduce variablies

a.CRIM - per capita crime rate by town
b.ZN - proportion of residential land zoned for lots over 25,000 sq.ft.
c.INDUS - proportion of non-retail business acres per town.
d.CHAS - Charles River dummy variable (1 if tract bounds river; 0 otherwise)
e.NOX - nitric oxides concentration (parts per 10 million)
f.RM - average number of rooms per dwelling

g.AGE - proportion of owner-occupied units built prior to 1940

h.DIS - weighted distances to five Boston employment centres

i.RAD - index of accessibility to radial highways

j.TAX - full-value property-tax rate per us 10,000

k.PTRATIO - pupil-teacher ratio by town

l.B - $1000(Bk - 0.63)^2$ where Bk is the proportion of blacks by town

m.LSTAT - lower status of the population

n.MEDV - Median value of owner-occupied homes in us 1000's

(2)Preprocess: 對input data 做normalization

III. Implement for Regression problem

A. *Model*

以下是3-layer MLP的model

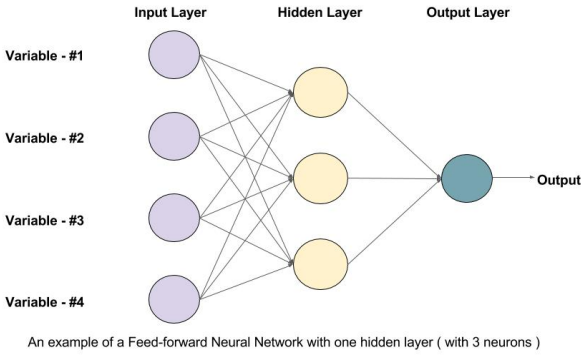

Fig. 1. 3-layer MLP

B. *Target*

$$min \quad Lossfunction = min \sum_i \left\| \hat{y}^i - y^i \right\|^2$$

C. *Results*

1. n hidden為30, epochs為1000, eta為0.002,minibatch size為6, train size和test size拆成7比3

result: train cost=24.4277, test cost= 8.1180

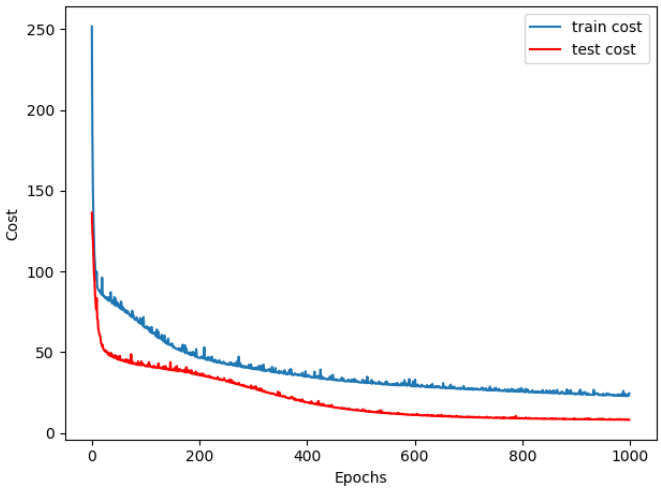

Fig. 2. result1

2. n hidden為50, epochs為1000, eta為0.002,minibatch size為6, train size和test size拆成7比3

result: train cost=20.5352, test cost= 16.3906

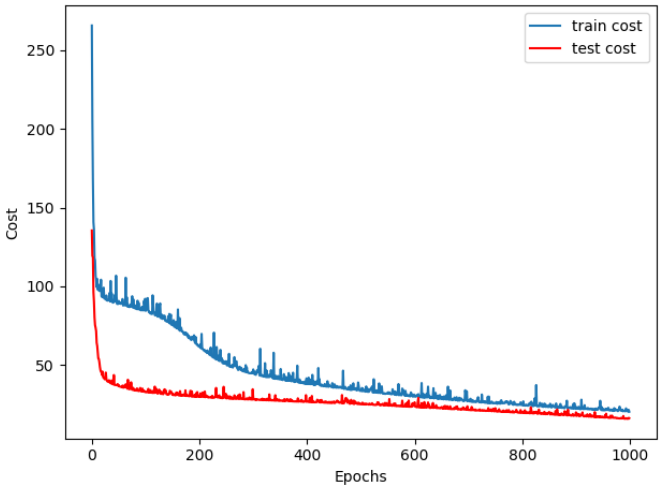

Fig. 3. result2