
2. Bishop Ex 4.15.

3. Write a computer program to classify the 3's and the 7's from the MNIST data using logistic regression. Get the data from http://www.cs.nyu.edu/~roweis/data/mnist_all.mat.

Use gradient descent using the expression 4.91 in Bishop. Monitor $E$ in each learning step and ensure that the learning rate $\eta$ is as large as possible, but small enough so that $E$ decreases in each step. Hint: Make sure that $0 < y < 1$ so that $\log y, \log(1 - y)$ well defined.

Here is some Matlab code to get you started.

```matlab
clear all
load mnistALL.mat
Xmnist=mnist.train_images;
ymnist=mnist.train_labels;

X3=Xmnist(:,:,find(ymnist==3)); % train images of 3's
X7=Xmnist(:,:,find(ymnist==7)); % train images of 7's
p3=size(X3,3); % number of train images of class 3
p7=size(X7,3); % number of train images of class 7

n=size(X3,1).^2; % input dimension
x3=reshape(X3,n,p3); % input data of 3's
x7=reshape(X7,n,p7); % input data of 7's
x3(n+1,:)=1; % extra input dimension for thresholds
x7(n+1,:)=1;
x=[x3,x7]';
t=zeros(1,p3+p7); % labels 0,1 for the two classes
t(1:p3)=0;
t(p3+1:p3+p7)=1;
x=double(x); % matlab does not like the mnist data format
p=size(x,1); % total number of samples
n=size(x,2); % total input dimension

w=zeros(1,n); % weights w(1:n-1) threshold w(n)
eta= % learning rate
dw_max=1; % max change in w in one learning step
i=0;
while dw_max>1e-5;
i=i+1;
    your gradient descent code
end;
```

fprintf('convergence after %d iteration: E = %f, dw_max = %f
',i,E,dw_max)

% classification performance on training set
x3=double(x3);
x7=double(x7);
error1=sum(w*x3>0);
error2=sum(w*x7<0);
fprintf('train set performance
');
fprintf('%d out of %d images of 3s misclassified
',error1,p3)
fprintf('%d out of %d images of 7s misclassified
',error2,p7)

% classification performance on training set
Xtest=mnist.test_images;
t=mnist.test_labels;
X3t=Xtest(:,:,find(ytest==3)); % test images of 3's
X7t=Xtest(:,:,find(ytest==7)); % test images of 7's
p3t=size(X3t,3); % number of images of class 3
p7t=size(X7t,3); % number of images of class 7

n=size(X3,1).^2;
x3t=reshape(X3t,n,p3t); % input data of 3's
x7t=reshape(X7t,n,p7t); % input data of 7's
x3t(n+1,:)=1; % extra input dimension for thresholds
x7t(n+1,:)=1;
x3t=double(x3t);
x7t=double(x7t);

error1=sum(w*x3t>0);
error2=sum(w*x7t<0);
fprintf('test set performance
');
fprintf('%d out of %d images of 3s misclassified
',error1,p3t)
fprintf('%d out of %d images of 7s misclassified
',error2,p7t)