Aditi Mallavarapu

CS 412 Homework 1

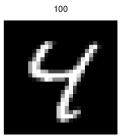
Problem 1 a)

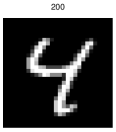
```
%% Classify the MNIST digits using a one nearest neighbour classifier and Euclidean distance
%% This file is modified from pmtk3.googlecode.com

load('mnistData');

% set training & testing
errorRate= [];
errorndx = 1;
varlimit = [100, 200,500,1000,2000,5000,10000]
for entry = varlimit

  trainndx = 1:entry;
  testndx = 1:10000;
  ntrain = length(trainndx);
  ntest = length(testndx);
  Xtrain = double(reshape(mnist.train_images(:,:,trainndx),28*28,ntrain)');
  Xtest  = double(reshape(mnist.test_images(:,:,testndx),28*28,ntest)');

   ytrain = (mnist.train_labels(trainndx));
   ytest  = (mnist.test_labels(testndx));

% Precompute sum of squares term for speed
   XtrainSOS = sum(Xtrain.^2,2);
   XtestSOS  = sum(Xtest.^2,2);
```

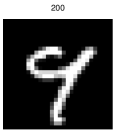
```
% fully solution takes too much memory so we will classify in batches
% nbatches must be an even divisor of ntest, increase if you run out of memory
   if ntest > 2000
     nbatches = 50;
   else
     nbatches = 5;
   end
   batches = mat2cell(1:ntest,1,(ntest/nbatches)*ones(1,nbatches));
   ypred = zeros(ntest,1);
   closestndx = [];
% Classify
   for i=1:nbatches
     dst = sqDistance(Xtest(batches{i},:),Xtrain,XtestSOS(batches{i},:),XtrainSOS);
     [junk,closest] = min(dst,[],2);
     ypred(batches{i}) = ytrain(closest);
     closestndx(batches{i}) = closest;
   end
% Report

   errorRate(errorndx) = mean(ypred ~= ytest);
   fprintf('Error Rate: %.2f%%\n',100*errorRate(errorndx));
   errorndx = errorndx + 1;
   imagesamp= [];
  %find the images that were misclassified
   imagesamp = (ypred~=ytest);

   for i = 1 : length(imagesamp)
     if (imagesamp(i) == 1)
       index = i;
```

```
      figure,
      subplot(2,1,1)
      imshow(mnist.test_images(:,:,index)) % the misclassified image
      title(entry)
      subplot(2,1,2)
      imshow(mnist.train_images(:,:,closestndx(index))) % the nearest neighbor image
      title(entry)
      break;
    endif
  end

 end
%%% Plot example

% line plot example random data
figure, plot(errorRate)
ylabel('accuracy')
```

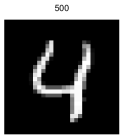
Misclassified Images plots (Upper image) and their nearest neighbor (lower image)

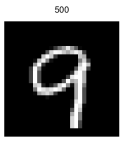




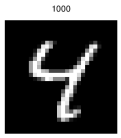


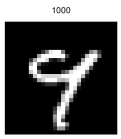
Training Size 100 *Training Size 200*



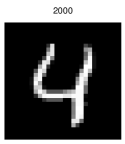


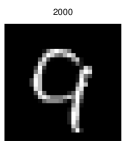
Training Size 500



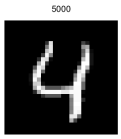


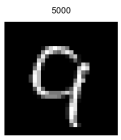
Training Size 1000



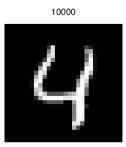


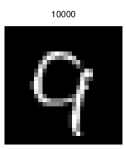
Training Size 2000



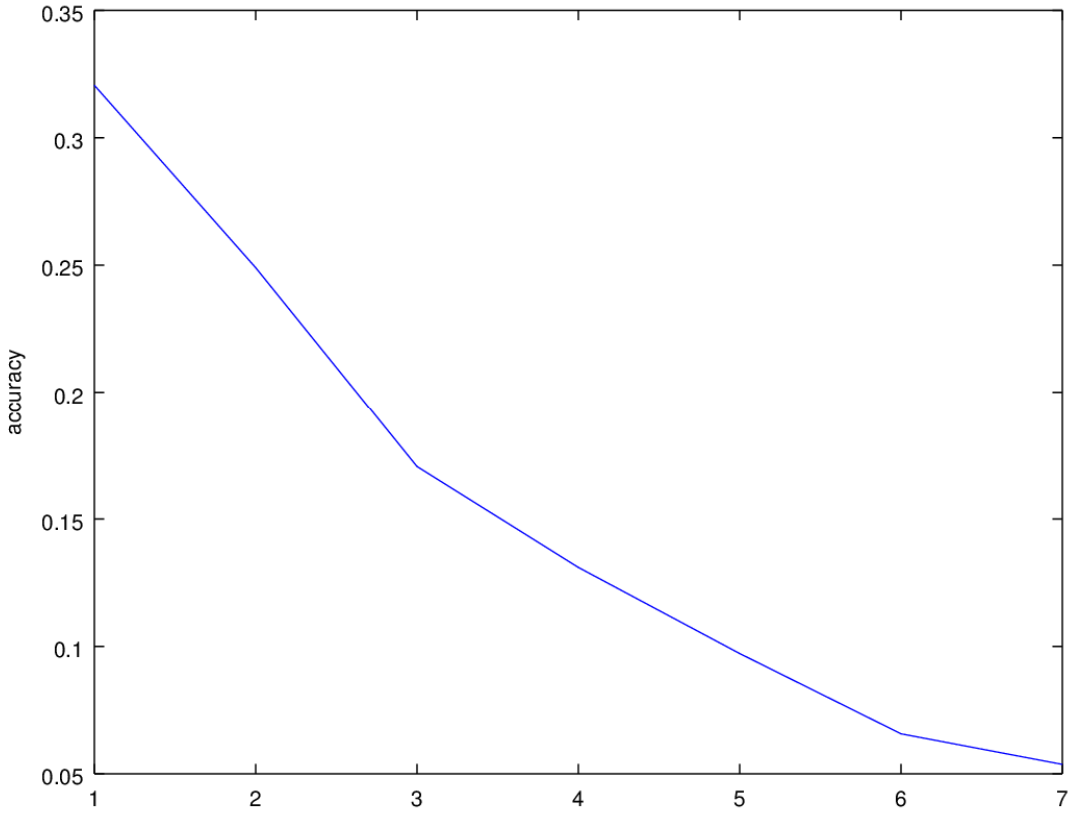


Training Size 5000





Training Size 10000



Error Plot