% Jiaqi Zhang

clear all; close all; clc

% Loading data

trainSet = loadMNISTImages('train-images-idx3-ubyte');

trainLabel = loadMNISTLabels('train-labels-idx1-ubyte');

testSet = loadMNISTImages('t10k-images-idx3-ubyte');

testLabel = loadMNISTLabels('t10k-labels-idx1-ubyte');

[trainDim, trainNum] = size(trainSet);

[testDim, testNum] = size(testSet);

%%

errorRate = [];

for totalRun = 1: 5

validError = [];

ite = 5; % Iterations for validation

trainSize = 50000;

validSize = trainNum - trainSize;

ASet = [];

%% Cross Validation

for count = 1: ite

% Pick random indices.

index = randperm(trainNum);

indexTrain = index(1:trainSize);

indexValid = index(trainSize+1: end);

% Seperating training/testing data

vTrainLabel = zeros(10, trainSize);

validLabel = zeros(10, validSize);

vTrainSet = zeros(trainDim,trainSize);

validSet = zeros(trainDim, validSize);

% Create label matrices.

for i = 1 : trainSize

vTrainLabel(trainLabel(indexTrain(i)) + 1,i) = 1;

vTrainSet(:,i) = trainSet(:, indexTrain(i));

end

for j = 1 : validSize

validLabel(trainLabel(indexValid(j)) + 1,j) = 1;

validSet(:,j) = trainSet(:, indexValid(j));

end

A = vTrainLabel \* pinv(vTrainSet); % A = S\* pinv(X)

vResult = A \* validSet;

[M,I] = max(vResult); % Maximum, index of Maximum

vResultLabel = zeros(10, validSize);

for k = 1 : validSize

vResultLabel(I(k),k) = 1;

end

vError = vResultLabel - validLabel;

error = nnz(vError)/2; % #non-zero = 2 \* mismatch.

validError = [validError; error/validSize];

ASet = [ASet; A];

end

plot(validError,'o');

title('Error Rate for Each Trial')

xlabel('Trials');

ylabel('Error Rate')

testLabelMatrix = zeros(10, testNum);

resultMatrix = zeros(10, testNum);

%Pick A that gives minimum error rate.

[M, I] = min(validError);

A = ASet((I-1)\*10+1:I\*10,:);

result = A\*testSet;

for k = 1 : testNum

[M,I] = max(result);

resultMatrix(I(k),k) = 1;

end

for i = 1 : testNum

testLabelMatrix(testLabel(i) + 1,i) = 1; % resulting matrix that has 1 at label\_value + 1 position, since value range from 0 to 9

end

errorMatrix = resultMatrix - testLabelMatrix;

errorCount = nnz(errorMatrix)/2;

errorRate = [errorRate; errorCount/testNum];

end

figure;

plot(errorRate, 'o'); %0.1473

title('Error Rate for Test Set');

xlabel('Run #');

ylabel('Error Rate');