Handwriting Recognition Using Bagged Classification Trees

[View all machine learning examples](https://www.mathworks.com/solutions/machine-learning/examples.html)

This example shows how to recognize handwritten digits using an ensemble of bagged classification trees. Images of handwritten digits are first used to train a single classification tree and then an ensemble of 200 decision trees. The classification performance of each is compared to one another using a confusion matrix.

Load Training and Test Data

See the references section for information on obtaining the dataset.

clear

load('usps\_all');

reduce\_dim = false;

X = double(reshape(data,256,11000)');

ylabel = [1:9 0];

y = reshape(repmat(ylabel,1100,1),11000,1);

clearvars data

Visualize Six Random Handwritten Samples

figure(1)

for ii = 1:6

subplot(2,3,ii)

rand\_num = randperm(11000,1);

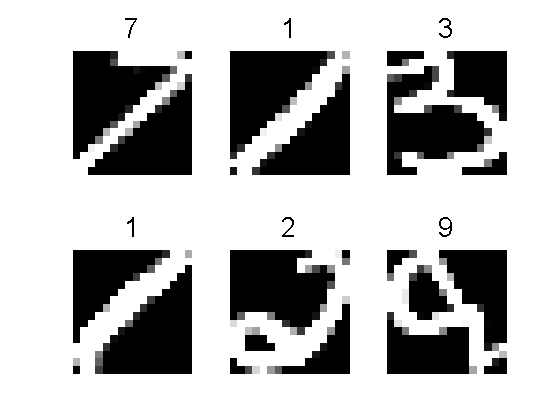
image(reshape(X(rand\_num,:),16,16))

title((y(rand\_num)),'FontSize',20)

axis off

end

colormap gray



Randomly Partition the Data into Training and Validation Sets

cv = cvpartition(y, 'holdout', .5);

Xtrain = X(cv.training,:);

Ytrain = y(cv.training,1);

Xtest = X(cv.test,:);

Ytest = y(cv.test,1);

Train and Predict Using a Single Classification Tree

mdl\_ctree = ClassificationTree.fit(Xtrain,Ytrain);

ypred = predict(mdl\_ctree,Xtest);

Confmat\_ctree = confusionmat(Ytest,ypred);

Train and Predict Using Bagged Decision Trees

mdl = fitensemble(Xtrain,Ytrain,'bag',200,'tree','type','Classification');

ypred = predict(mdl,Xtest);

Confmat\_bag = confusionmat(Ytest,ypred);

Compare Confusion Matrices

figure,

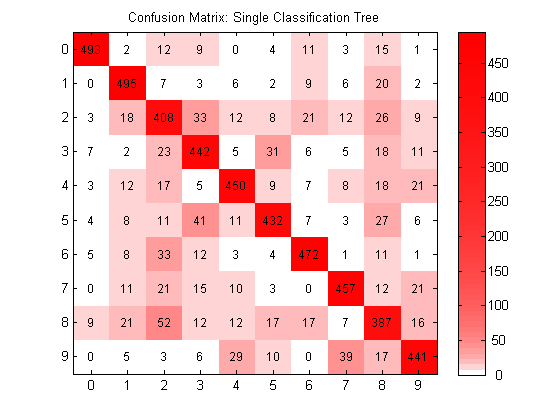
heatmap(Confmat\_ctree, 0:9, 0:9, 1,'Colormap','red','ShowAllTicks',1,'UseLogColorMap',true,'Colorbar',true);

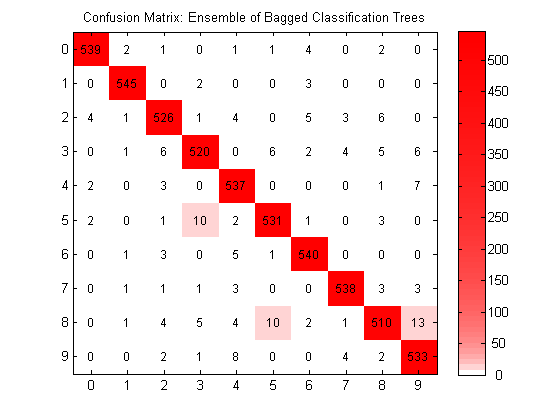
title('Confusion Matrix: Single Classification Tree')

figure,

heatmap(Confmat\_bag, 0:9, 0:9, 1,'Colormap','red','ShowAllTicks',1,'UseLogColorMap',true,'Colorbar',true);

title('Confusion Matrix: Ensemble of Bagged Classification Trees')





Bagged classification trees perform much better than a single classification tree on the training set since the confusion matrix is more dominantly diagonal.

Visualization generated using [Customizable Heat Maps](https://www.mathworks.com/matlabcentral/fileexchange/24253-customizable-heat-maps).

Reference and License

MAT file for the images are located here: [cs.nyu.edu/~roweis/data.html](http://www.cs.nyu.edu/~roweis/data.html).