Table of Contents

Splitting my training data into train and validation sets

```
load('training_data.mat')
load('test_data.mat')
load('column_names.mat')

training_data = data(:,1:14);
training_labels = data(:,15);

pick = 0.9*32724;
pick = double(pick);
validation_index = randsample(32724,pick);

val_label = training_labels(validation_index);
val_matrix = training_data(validation_index,:);
rest = setdiff(1:5172,validation_index);
train_matrix = training_data(rest,:);
train_label = training_labels(rest)';
train_label = double(train_label);
```

```
Warning: Integer operands are required for colon operator when used as
 index
```

Making a tree

```
test_matrix = test_data;

%dt = DecisionTree();
%trained_tree = dt.train(train_matrix,train_label);
pred_label = zeros(1,length(test_data));
for i = 1:length(test_matrix)
testing_data = test_matrix(i,:);
pred_label(i) = trained_tree.predict(testing_data,column_names);
end
```

```
Undefined variable "trained_tree" or class "trained_tree.predict".

Error in HW05 (line 29)
pred_label(i) = trained_tree.predict(testing_data,column_names);
```

Validation Testing

```
%accuracy = sum(pred_label==val_label)/length(val_label);
%display(accuracy,'The Accuracy is');
```

Spam test

```
Submission_matrix = [(1:length(test_data))', pred_label'];
```

Random Forests

```
%training_labels = double(training_labels');
pred_label = zeros(length(val_label),1);
for index = 1:20
    [data, label] = bagging(train_matrix,train_label');
    df = DecisionTree();
    classifier = df.train(data,label);
    pred_label = zeros(1,length(val_label));
    for i = 1:length(val_matrix)
        testing_data = val_matrix(i,:);
        pred_label(i,index) =
 classifier.predict(testing_data,column_names);
    end
end

predictions = mode(pred_label');
```

Submission

```
Submission_matrix_f = [(1:length(test_data))', predictions'];
%accuracy = 100*sum(predictions==val_label)/length(val_label);
%display(accuracy,'The Accuracy is (in %)');
```

Published with MATLAB® R2015b