Table of Contents

Course Project Assingment

```
        Main Porject .m File Created by Akash Mukesh Joshi - USC ID - 4703642421
```

```
tic; clc; clear; close all; load('diabeticdataset.mat');
```

Data Separation

```
[feature_train,label_train,feature_test,label_test_um] =
 dataseparation(diabeticdata);
```

Preprocessing

```
feature_names = diabeticdata.Properties.VariableNames;
 feature_names(:,end) = [];
[feature_train_pp,feature_names,label_train,~] =
 preprocessing(feature_train,feature_names,label_train,0);
feature_names = diabeticdata.Properties.VariableNames;
 feature_names(:,end) = [];
[feature_test_pp,feature_names,label_test,rowdh] =
 preprocessing(feature_test,feature_names,label_test_um,1);
label_shift = label_test_um(rowdh(:),:);
predicted_shift = ones(size(rowdh,1),1);
```

Classification Learner Data Preparation

```
xtrain = [feature_train_pp label_train];
```

Equal Priors

```
[~,I] = sort(label_train);
```

```
feature_train_pp = feature_train_pp(I,:);
label_train = label_train(I,:);

unique(label_train);
cvk = (histc(label_train, unique(label_train)));
cvk(1) = ceil(cvk(1)/cvk(3));
cvk(2) = ceil(cvk(2)/cvk(3));
cvk(3) = 1;

feature_train_pp =
 [feature_train_pp(1:3903,:);feature_train_pp(22887:26789,:) ;feature_train_pp(362
label_train =
 [label_train(1:3903,:);label_train(22887:26789,:) ;label_train(36210:end,:)];
```

Regression Models

```
[~, ~, ~, ~, c, ~] = lars(feature_train_pp,
 double(label_train), 'lar', [],[]);
[new_patterns, new_targets, pattern_numbers] =
 Sequential_Feature_Selection(feature_train_pp,
 double(label_train'), '[''Forward'',2, ''LS'', []]');
feature_train_pp =
 pca(feature_train_pp','Algorithm','svd','NumComponents',2);
feature_test_pp =
 pca(feature_test_pp','Algorithm','svd','NumComponents',2);
```

Correlation Matrix

```
correl = corrcoef(feature_train_pp,'rows','pairwise');
```

Histogram Polt

```
close all;
for i = 1:34
    figure; histogram(feature_train_pp(:,i));
end
```

Plot Data Set

```
pt = [14 15];
data = feature_train_pp(:,pt);
figure; hold on;
plot(data(label_train == 1,1),data(label_train == 1,2), 'ro');
plot(data(label_train == 2,1),data(label_train == 2,2), 'gd');
plot(data(label_train == 3,1),data(label_train == 3,2), 'b*');
```

SVM Model

```
mod = svmtrain(double(label_train), feature_train_pp(:,ft),'-b 1 -t 0
 -c 100 -m 10000 -q');
```

```
[predicted_label,accuracy,~] = svmpredict(double(label_test),
 feature_test_pp(:,ft), mod, '-q');
[m_f1,~] =
 classification_report(double(label_test),double(predicted_label'),
 1);
```

MSE Model

```
predicted_label =
 multiclass(feature_train_pp',double(label_train'),feature_test_pp', '[''OAA'',
 0, ''Perceptron'', 2000]');
disp(mean(predicted_label == double(label_test')));
[m_f1,~] =
 classification_report(double(label_test),double(predicted_label'),
 1);
```

Ensemble Subspace KNN Matlab Toolbox

Finding the best value of nearest neighbors for the KNN Classifier

```
[N,D] = size(feature_train_pp);
K = round(logspace(0,log10(N),10)); % number of neighbors
cvloss = zeros(numel(K),1);
for k=1:numel(K)
    knn = fitcknn(feature_train_pp,label_train,...
        'NumNeighbors',K(k),'CrossVal','On');
    cvloss(k) = kfoldLoss(knn);
end
figure; % Plot the accuracy versus k
semilogx(K,cvloss);
xlabel('Number of nearest neighbors');
ylabel('10 fold classification error');
title('k-NN classification');

% Findind the best subspace demesion value
NPredToSample = round(linspace(1,D,10)); % linear spacing of
 dimensions
cvloss = zeros(numel(NPredToSample),1);
learner = templateKNN('NumNeighbors',104);
for npred=1:numel(NPredToSample)
    subspace =
 fitensemble(feature_train_pp,label_train,'Subspace',100,learner,...
        'NPredToSample',NPredToSample(npred),'CrossVal','On');
    cvloss(npred) = kfoldLoss(subspace);
    fprintf('Random Subspace %i done.\n',npred);
end
figure; % plot the accuracy versus dimension
plot(NPredToSample,cvloss);
xlabel('Number of predictors selected at random');
ylabel('10 fold classification error');
title('k-NN classification with Random Subspace');

% Finding the best number of learners
```

```
ens =
 fitensemble(feature_train_pp,label_train,'Subspace',100,learner,...
    'NPredToSample',104,'CrossVal','on');
plot(kfoldLoss(ens,'Mode','Cumulative'))
xlabel('Number of learners in ensemble');
ylabel('10 fold classification error');
title('k-NN classification with Random Subspace');

% Finding the best K Fold value
for k=10:10:50
    subspace =
 fitensemble(feature_train_pp,label_train,'Subspace',10,learner,...
        'NPredToSample',30,'CrossVal','On','KFold',k);
    cvloss(k) = kfoldLoss(subspace)
end
plot(10:10:50,cvloss);
xlabel('Number of K Folds');
ylabel('Classification error');
title('k-NN classification with Random Subspace');

[trainedClassifier, validationAccuracy] =
 trainClassifier([feature_train_pp label_train]);
predicted_label = trainedClassifier.predictFcn(feature_test_pp);
[m_f1,~] = classification_report([label_test],[predicted_label], 1);
```

Naive Bayes Model

```
Mdl =
 fitcnb(feature_train_pp,double(label_train),'Distribution','normal','CrossVal','o
for i = 1 : 10
    pred_label(:,i) = predict(Mdl.Trained{i,1},feature_test_pp);
end
predicted_label = mode(pred_label,2);
[m_f1,~] = classification_report([label_test],[predicted_label], 1);
```

Confusion Matrix

```
[m_f1,~] =
 classification_report(double(label_test),double(predicted_label'),
 1);
```

Published with MATLAB® R2016a