# ICT4Health Lab 2 - Principal component regression (PCR)

## Table of Contents

In this lab, regression is performed with same data as Lab 1 on following features, I. the total UPDRS value (column 7), II. the motor UPDRS (column 5), but with only L uncorrelated features instead of all features measured. These uncorrelated features are extracted using Principle component analysis (PCA)

# Data Preparation

```
close all;
clear all;
clc;

% loading data saved in previous lab
load('parkinsonsdat_for_regression.mat','data');
[nrows, ncols] = size(data);

%Selecting training set data range from given data
for i=1:nrows
    if data(i,1) > 36       %Data until patient number 36 is selected
        break
    end
    row_cut = i;
end
data_train = data(1:row_cut, :);
data_test = data(row_cut+1:end,:);

%Standardising training and testing set data
m_data_train = mean(data_train,1);
v_data_train = std(data_train,1);
data_train_norm = (data_train - m_data_train)./v_data_train;

m_data_test = mean(data_test,1);
v_data_test = std(data_test,1);
data_test_norm = (data_test - m_data_test)./v_data_test;

%Testing and training data are standardised after splitting
%to avoid the inclusion of future information while developing model
```

# PCR for feature F0 = 7

```
F0 = 7;
```

```matlab
% selecting testing and training data sets
Y_train = data_train_norm(:,F0);
X_train = data_train_norm;
X_train(:,F0) = [];

Y_test = data_test_norm(:,F0);
X_test = data_test_norm;
X_test(:,F0) = [];

% Principal component analysis (PCA)
[N,F] = size(X_train);
R = (X_train'*X_train)/N; % Covariance matrix
[U,V] = eig(R);
Z = X_train*U;

Zn = (Z*(V^-0.5))/sqrt(N);
Zy = Zn'*Y_train;
yhat_train = Zn*Zy;
a = inv(X_train'*X_train)*X_train'*Y_train;

% removing features with low eigen values i.e. highly correlated
 features
% hence, keeping only important uncorrelated features for regression
d=diag(V);d1=d/sum(d);d1c=cumsum(d1);
removed_eigen=1e-3;
nrem=(d1c<removed_eigen);
UL=U;UL(:,nrem)=[];
o = ones(N,1);
z=X_train*UL;
%ZnL=z./(o*sqrt(std(z)));

VL = V; VL(:,nrem) = []; VL(nrem,:) = [];
ZnL = ((X_train*UL)*(VL^(-0.5)))/sqrt(N);
ZyL = ZnL'*Y_train;
yhat_train_L = ZnL*ZyL;

aL = (1/N)*UL*(VL^(-0.5))*UL'*X_train'*Y_train;

yhat_test = X_test*a;
yhat_test_L = X_test*aL;

figure
subplot(2,2,1)
plot(Y_train,yhat_train,'b*')
title('$Y_{train} \, vs \, \hat{Y}_{train}$','Interpreter','latex')
subplot(2,2,2)
plot(Y_train,yhat_train_L,'g*')
title('$Y_{train} \, vs \, \hat{Y}_{train-
PCR}$','Interpreter','latex')
subplot(2,2,3)
plot(Y_test,yhat_test,'b*')
title('$Y_{test} \, vs \, \hat{Y}_{test}$','Interpreter','latex')
subplot(2,2,4)
plot(Y_test,yhat_test_L,'g*')
```
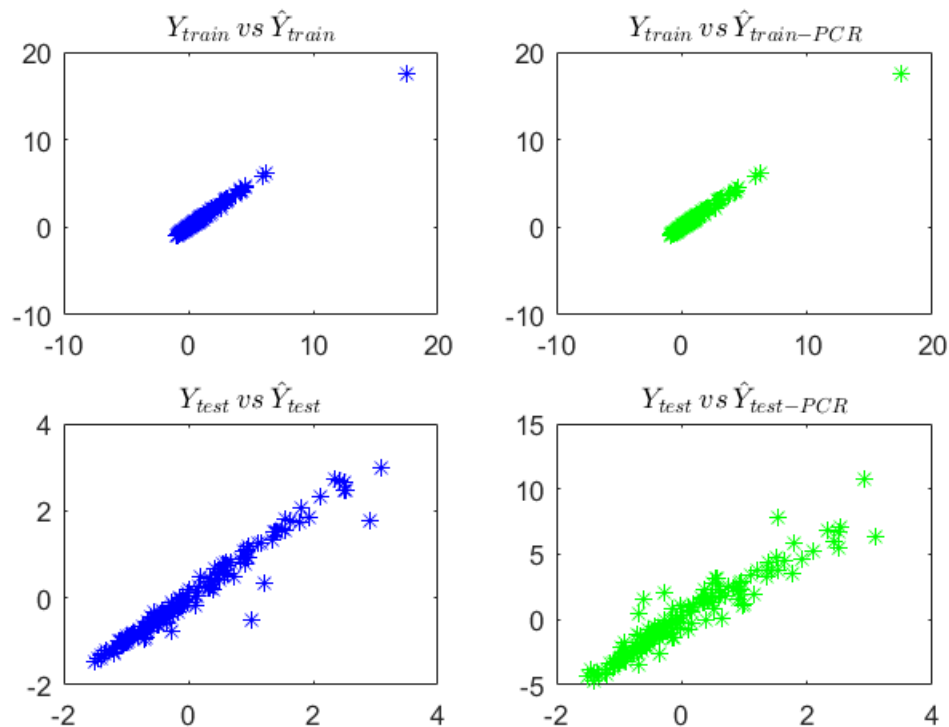
```
title('$Y_{test} \, vs \, \hat{Y}_{test-PCR}$','Interpreter','latex')
suptitle('For F0=7')

figure
nbins = 50;
subplot(2,2,1)
histogram(Y_train-yhat_train,nbins);
title('$Y_{train} \, - \, \hat{Y}_{train} \,
\, ,nbins=50$','Interpreter','latex')
axis([-1 1 0 155])
subplot(2,2,2)
histogram(Y_train-yhat_train_L,nbins);
title('$Y_{train} \, - \, \hat{Y}_{train}^{L} \,
\, ,nbins=50$','Interpreter','latex')
axis([-1 1 0 155])
subplot(2,2,3)
histogram(Y_test-yhat_test,nbins);
title('$Y_{test} \, - \, \hat{Y}_{test} \,
\, ,nbins=50$','Interpreter','latex')
axis([-1 1 0 155])
subplot(2,2,4)
histogram(Y_test-yhat_test_L,nbins);
title('$Y_{test} \, - \, \hat{Y}_{test}^{L} \,
\, ,nbins=50$','Interpreter','latex')
suptitle('For F0=7')

figure
plot(1:21,a,1:21,aL);
title('$Weights for each features (F0=7)$','Interpreter','latex')
xlabel('$No.\,of\,features$','Interpreter','latex')
ylabel('$Weight$','Interpreter','latex')
legend('without PCA','with PCA','Location','southwest')
```
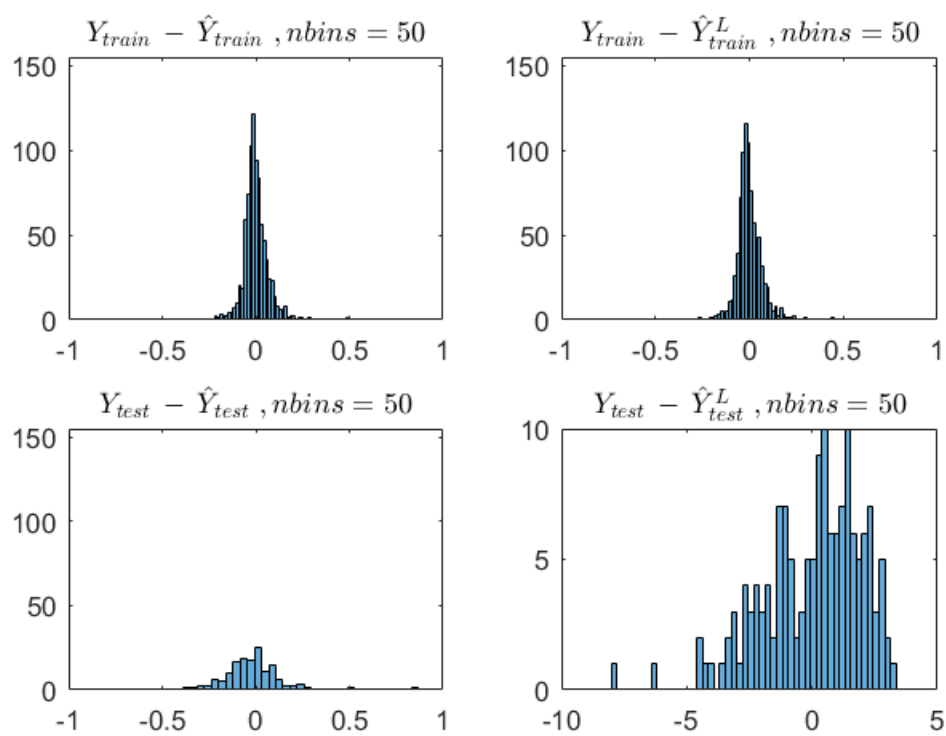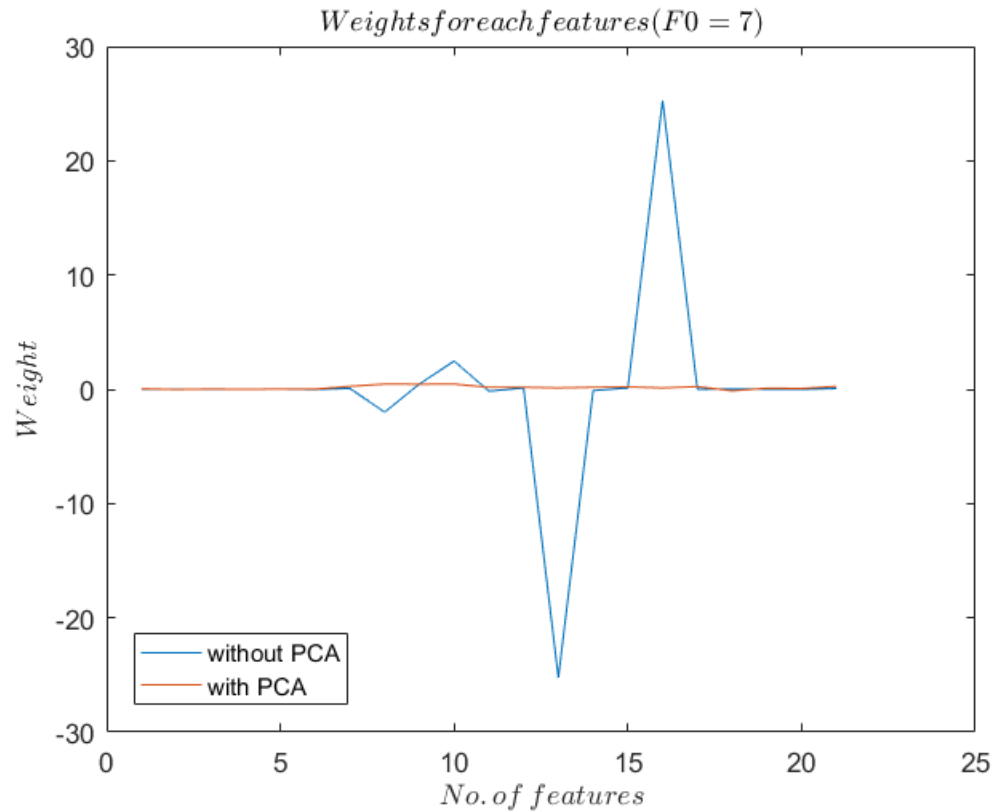
## For F0=7

$Y_{train}\ vs\ \hat{Y}_{train}$

$Y_{train}\ vs\ \hat{Y}_{train-PCR}$

$Y_{test}\ vs\ \hat{Y}_{test}$

$Y_{test}\ vs\ \hat{Y}_{test-PCR}$

## For F0=7

$Y_{train} - \hat{Y}_{train}\ , nbins = 50$

$Y_{train} - \hat{Y}_{train}^{L}\ , nbins = 50$

$Y_{test} - \hat{Y}_{test}\ , nbins = 50$

$Y_{test} - \hat{Y}_{test}^{L}\ , nbins = 50$

## PCR for feature F0 = 5

```
F0 = 5;
Y_train = data_train_norm(:,F0);
X_train = data_train_norm;
X_train(:,F0) = [];

Y_test = data_test_norm(:,F0);
X_test = data_test_norm;
X_test(:,F0) = [];

% Principal component analysis (PCA)

[N,F] = size(X_train);
R = (X_train'*X_train)/N; % Covariance matrix
[U,V] = eig(R);
Z = X_train*U;

Zn = (Z*(V^-0.5))/sqrt(N);
Zy = Zn'*Y_train;
yhat_train = Zn*Zy;
a = inv(X_train'*X_train)*X_train'*Y_train;

d=diag(V);d1=d/sum(d);d1c=cumsum(d1);
removed_eigen=1e-3;
```

```matlab
nrem=(d1c<removed_eigen);
UL=U;UL(:,nrem)=[];
o = ones(N,1);
z=X_train*UL;
%ZnL=z./(o*sqrt(std(z)));

VL = V; VL(:,nrem) = []; VL(nrem,:) = [];
ZnL = ((X_train*UL)*(VL^(-0.5)))/sqrt(N);
ZyL = ZnL'*Y_train;
yhat_train_L = ZnL*ZyL;

aL = (1/N)*UL*(VL^(-0.5))*UL'*X_train'*Y_train;

yhat_test = X_test*a;
yhat_test_L = X_test*aL;

figure
subplot(2,2,1)
plot(Y_train,yhat_train,'b*')
title('$Y_{train} \, vs \, \hat{Y}_{train}$','Interpreter','latex')
subplot(2,2,2)
plot(Y_train,yhat_train_L,'g*')
title('$Y_{train} \, vs \, \hat{Y}_{train-
PCR}$','Interpreter','latex')
subplot(2,2,3)
plot(Y_test,yhat_test,'b*')
title('$Y_{test} \, vs \, \hat{Y}_{test}$','Interpreter','latex')
subplot(2,2,4)
plot(Y_test,yhat_test_L,'g*')
title('$Y_{test} \, vs \, \hat{Y}_{test-PCR}$','Interpreter','latex')
suptitle('For F0=5')

figure
nbins = 50;
subplot(2,2,1)
histogram(Y_train-yhat_train,nbins);
title('$Y_{train} \, - \, \hat{Y}_{train} \,
\, ,nbins=50$','Interpreter','latex')
axis([-1 1 0 155])
subplot(2,2,2)
histogram(Y_train-yhat_train_L,nbins);
title('$Y_{train} \, - \, \hat{Y}_{train}^{L} \,
\, ,nbins=50$','Interpreter','latex')
axis([-1 1 0 155])
subplot(2,2,3)
histogram(Y_test-yhat_test,nbins);
title('$Y_{test} \, - \, \hat{Y}_{test} \,
\, ,nbins=50$','Interpreter','latex')
axis([-1 1 0 155])
subplot(2,2,4)
histogram(Y_test-yhat_test_L,nbins);
title('$Y_{test} \, - \, \hat{Y}_{test}^{L} \,
\, ,nbins=50$','Interpreter','latex')
suptitle('For F0=5')
```
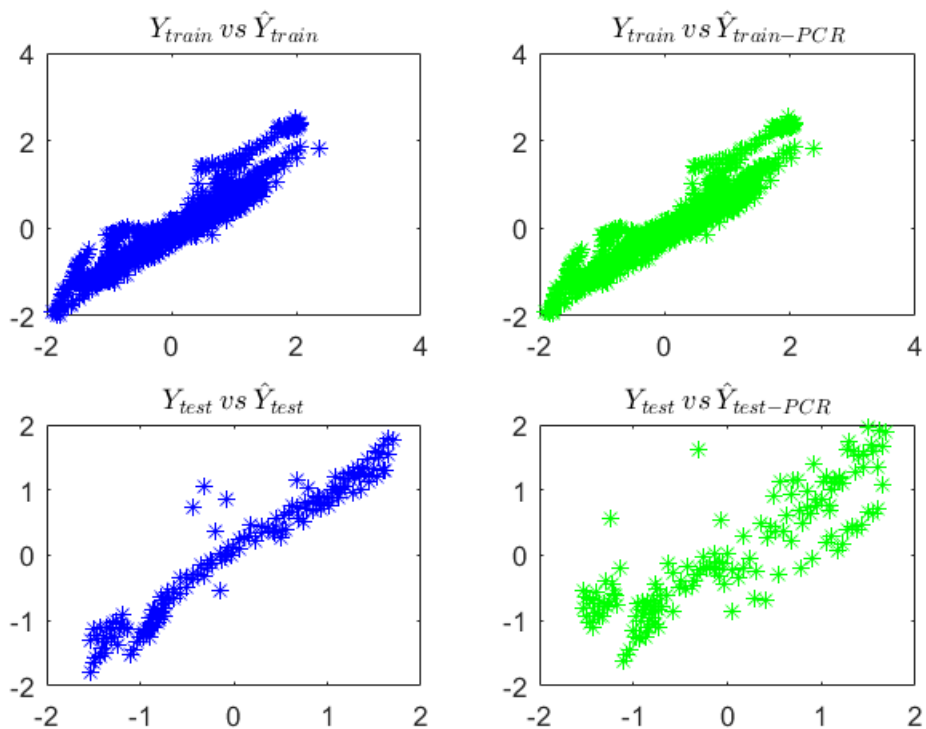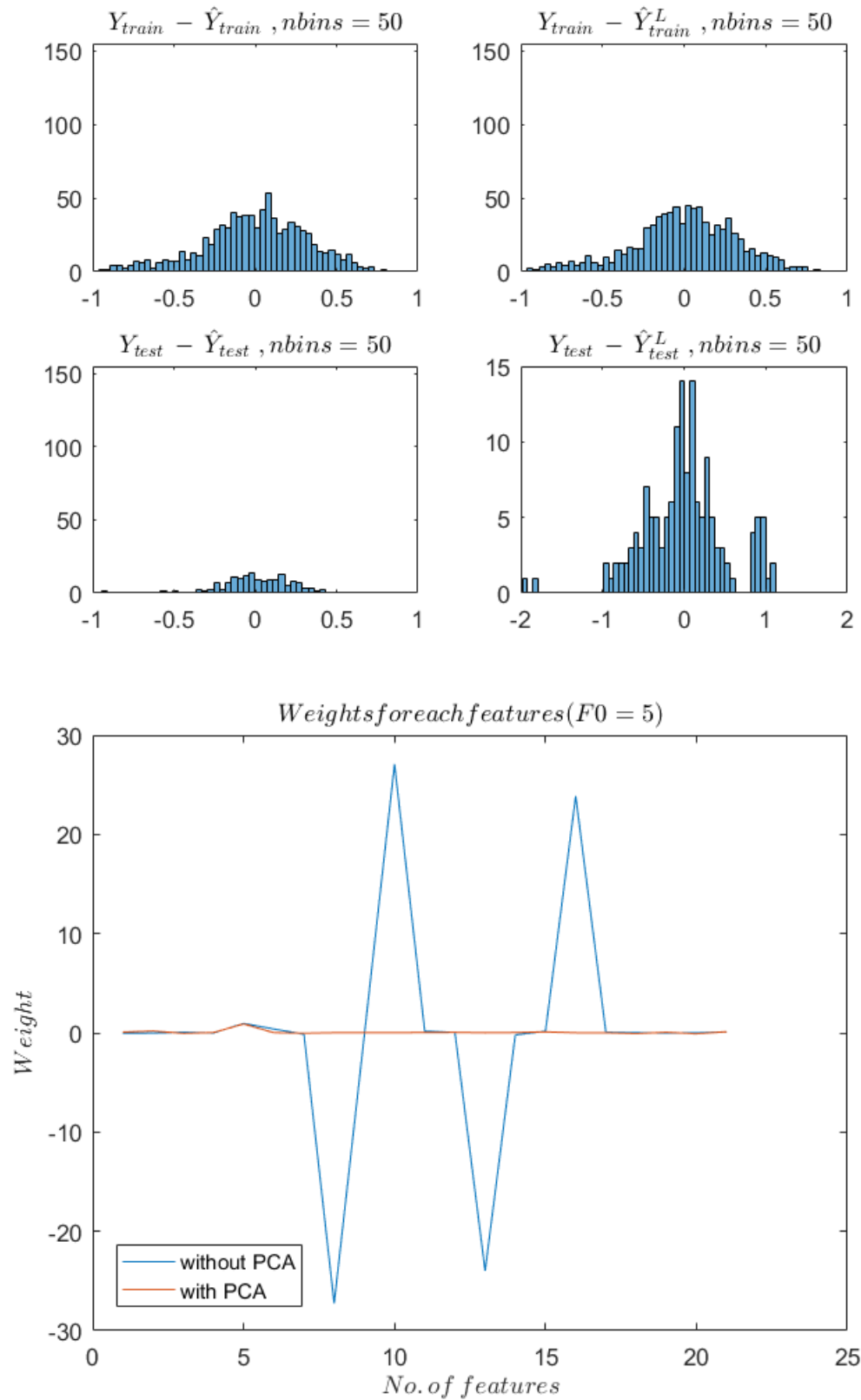
```
figure
plot(1:21,a,1:21,aL);
title('$Weights for each features (F0=5)$','Interpreter','latex')
xlabel('$No.\,of\,features$','Interpreter','latex')
ylabel('$Weight$','Interpreter','latex')
legend('without PCA','with PCA','Location','southwest')
```



For F0=5

## For F0=5

*Published with MATLAB® R2016b*