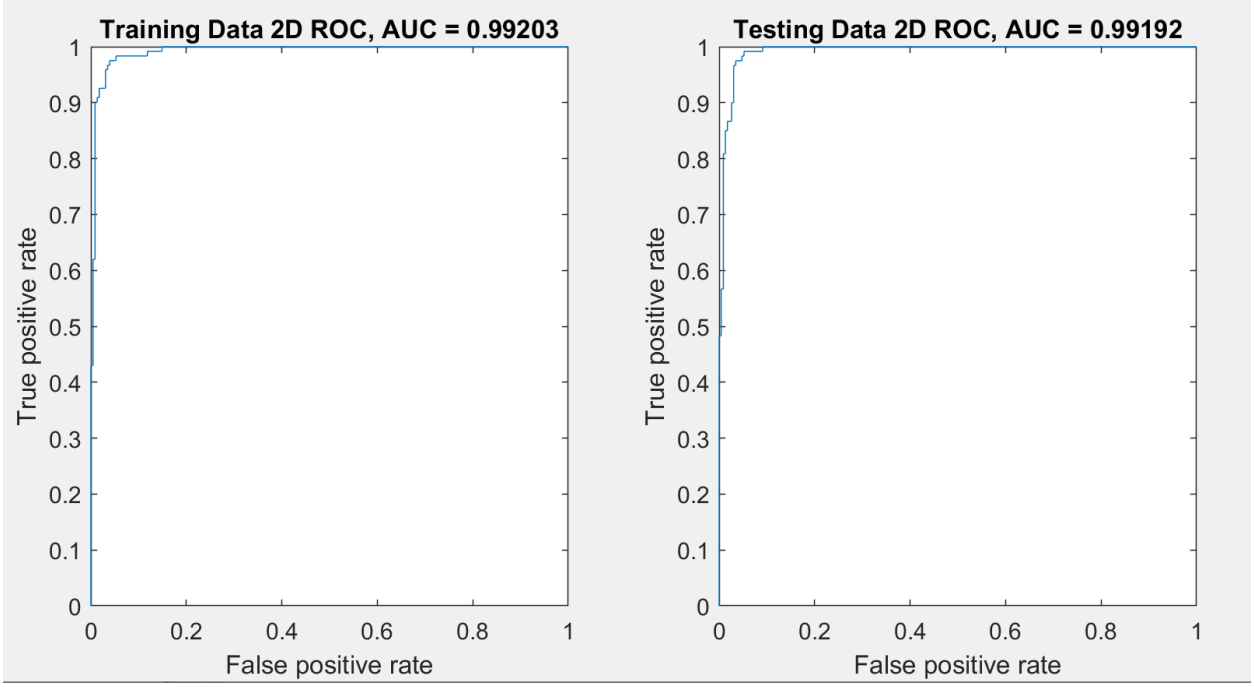
#1



Observations:

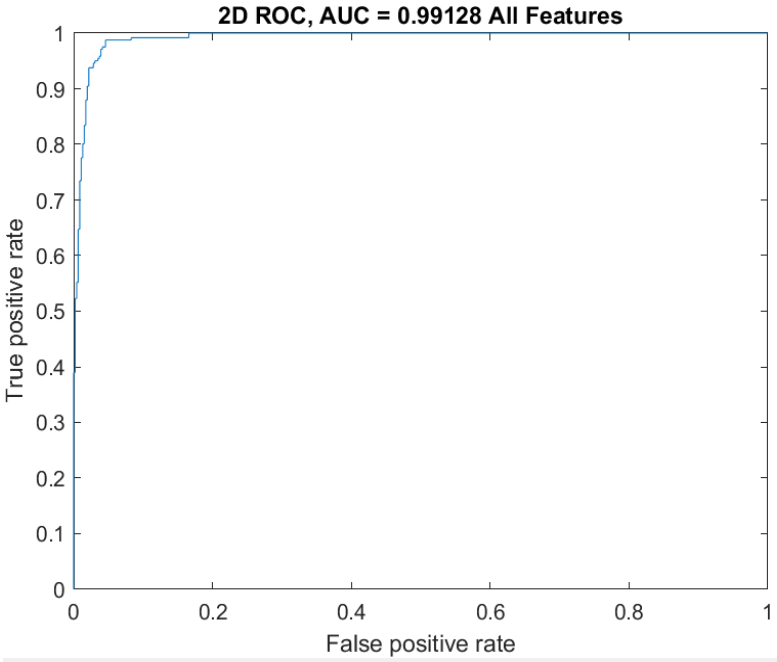
The training AUC is high at 99%, indicating a good fit of the linear model to the training data. The testing AUC is lower than the training AUC, indicating some degree of overfitting, but still reasonably high, indicating good generalization performance.

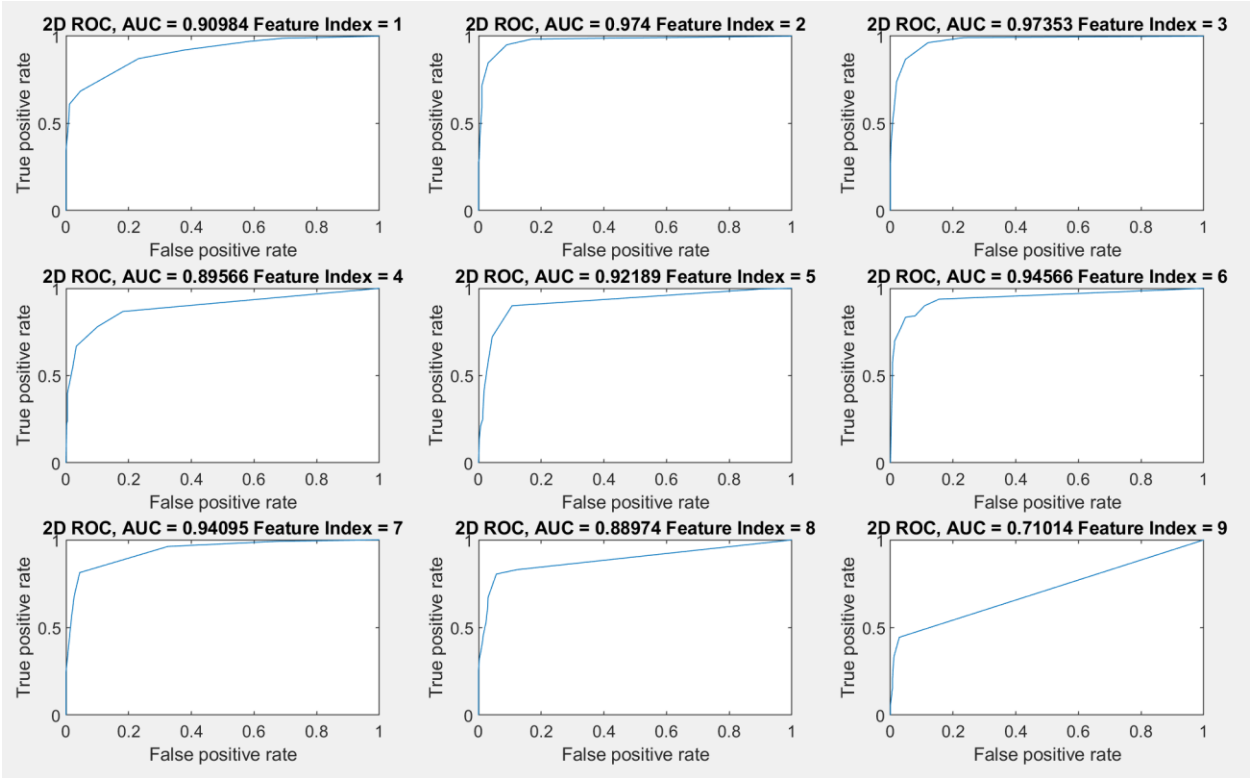
Code:

```
% Load cancer dataset
%
% Load cancer dataset
%
% Type help cancer_dataset for more info.
[X, d] = cancer_dataset;

% Split data into training and testing sets
n = length(X(1,:)); % number of examples in the dataset
train_idx = datasample(1:n, floor(n/2), 'Replace', false); % randomly
select floor(n/2) examples for training without replacement
test_idx = setdiff(1:n, train_idx); % use the remaining examples for
testing
X_train = X(:,train_idx); % input features for training examples
d_train = d(2,train_idx)'; % target variable for training examples
X_test = X(:,test_idx); % input features for testing examples
d_test = d(2,test_idx)'; % target variable for testing examples

% Training/MSE linear model creation
%
% "'" is transposing the variable
% and "\" is performing matrix left division.
```

```
w = X_train'\d_train; % compute weights that minimize mean squared
error

% Activation/testing
y_train = X_train'*w; % compute model outputs for training examples
y_test = X_test'*w; % compute model outputs for testing examples

% Compute 2D ROC curve for train data
[X2D_train,Y2D_train,T2D_train,AUC2D_train] =
perfcurve(d_train,y_train,1); % compute FPR, TPR, threshold, and AUC
for the 2D ROC curve

% Compute 2D ROC curve for test data
[X2D_test,Y2D_test,T2D_test,AUC2D_test] = perfcurve(d_test,y_test,1);
% compute FPR, TPR, threshold, and AUC for the 2D ROC curve

% Display results
disp(['Training AUC: ' num2str(AUC2D_train)]);
disp(['Testing AUC: ' num2str(AUC2D_test)]);

% Create a figure with two subplots
fig = figure;
subplot(1,2,1)

% Plot Training 2D ROC curve
plot(X2D_train,Y2D_train)
xlabel('False positive rate')
ylabel('True positive rate')
title(['Training Data 2D ROC, AUC = ' num2str(AUC2D_train)])

subplot(1,2,2)

% Plot 2D ROC curve
plot(X2D_test,Y2D_test)
xlabel('False positive rate')
ylabel('True positive rate')
title(['Testing Data 2D ROC, AUC = ' num2str(AUC2D_test)])
```

#2





observation:

Based on the ROC observations obtained from the code, it appears that some subsets of input variables perform better than others in predicting the target variable. The full 9-dimensional input yielded an AUC of around 99%, which is considered to be a good predictive performance. While most of the AUC's were around 90%, input variable 9 performed significantly worse, with a low AUC of 71%. This suggests that not all input variables may be equally important in predicting the target variable, and that a reduced input space dimensionality may indeed perform better.

code:

```
% Load cancer dataset
%
% Type help cancer_dataset for more info.
[X, d] = cancer_dataset;
% Training/MSE linear model creation all 9 features included
%
% "'" is transposing the variable
% and "\" is performing matrix left division.
% currently all 9 features are used in this code
w = X'\d(2,:)';
% Activation/testing
y = X'*w;
% Compute 2D ROC curve
%
% X2D,Y2D,T2D,AUC2D  are the output arguments that store the values of the
% FPR, the TPR, the threshold for the 2D ROC curve, and the AUC respectively.
[X2D,Y2D,T2D,AUC2D] = perfcurve(d(2,:),y',1);
% Plot 2D ROC curve
figure(1)
plot(X2D,Y2D)
xlabel('False positive rate')
ylabel('True positive rate')
title(['2D ROC, AUC = ' num2str(AUC2D), ' All Features'])
number_of_features = length(X(:,9));
%comparing the auc when only one feature is used for all 9 features
figure(2)
for i = 1:number_of_features
    % % generating an upper and lower bound if a random subset was desired
    % disp("the features include for this subet are: ")
    % lower_bound = round(8*rand(1))
    % disp("though")
    % upper_bound = round(9*rand(1))
    % Training/MSE linear model creation for each feature 1 through 9
    %
    % "'" is transposing the variable
    % and "\" is performing matrix left division.
    % currently all 9 features are used in this code
    w = X(i,:)'\d(2,:)';
    % Activation/testing
    y = X(i,:)'*w;
    % Compute 2D ROC curve
    %
    % X2D,Y2D,T2D,AUC2D  are the output arguments that store the values of
the
    % FPR, the TPR, the threshold for the 2D ROC curve, and the AUC
respectively.
    [X2D,Y2D,T2D,AUC2D] = perfcurve(d(2,:),y',1);
    % Plot 2D ROC curve
    %figure(i+1)
    % Set up subplot
    subplot(3, 3, i);
    plot(X2D,Y2D)
```

```
    xlabel('False positive rate')
    ylabel('True positive rate')
    title(['2D ROC, AUC = ' num2str(AUC2D), ' Feature Index = ', num2str(i)])
end
```