Instituto Superior Técnico

Departamento de Engenharia Electrotécnica e de Computadores

Machine Learning

1st Lab Assignment

Shift: Tuesday Group Number: 2

Number 81398 Name Maria Carolina Varandas Roque

Number 81013 Name José António Costa Coelho

Linear Regression

Linear Regression is a simple technique for predicting a real output *y* given an input $\mathbf{x}=(x_1, x_2,\ldots, x_P,)$ via the linear model

$$= \beta + \quad \beta$$

Typically there is a set of training data $T=\{(\mathbf{x}^i, y^i), i=1,.., N\}$ from which to estimate the coefficients $\beta=[\beta_0, \beta_1, \ldots,\beta_P]^T$. The Least Squares (LS) approach finds these coefficients by minimizing the sum of squares error

$$= -$$

The linear model is limited because the output is a linear function of the input variables x^k. However, it can easily be extended to more complex models by considering linear combinations of nonlinear functions, φ , of the input variables

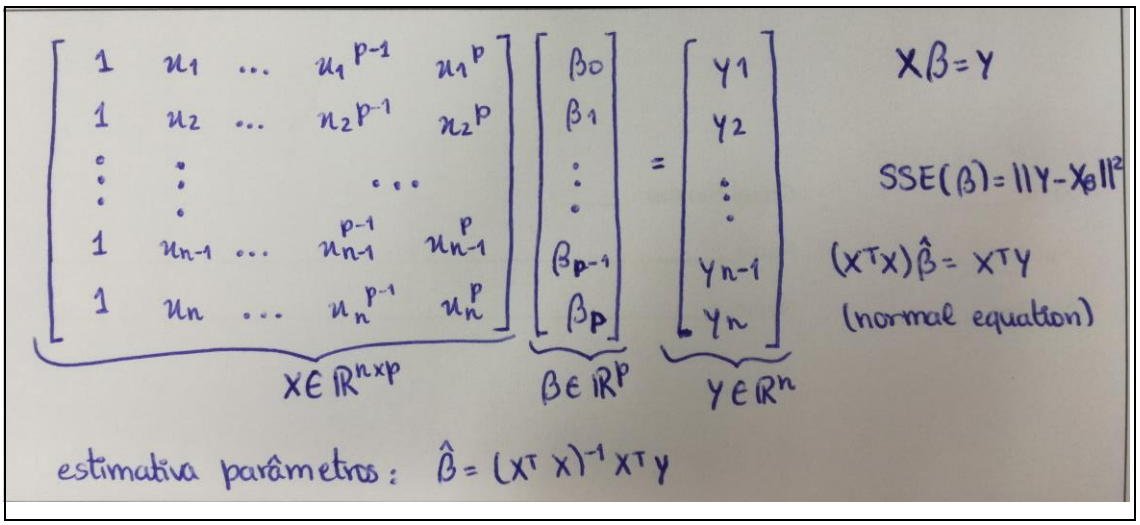
$$= \beta + \quad \beta \varphi$$

In this case the model is still linear in the parameters although it is nonlinear in **x**. Examples of nonlinear function include polynomial functions and Radial basis functions.

This assignment aims at illustrating Linear Regression. In the first part, we'll experiment linear and polynomial models. In the second part, we'll illustrate regularized Least Squares Regression. The second part of this assignment requires MatLab's Statistics Toolbox.

1. Least Squares Fitting

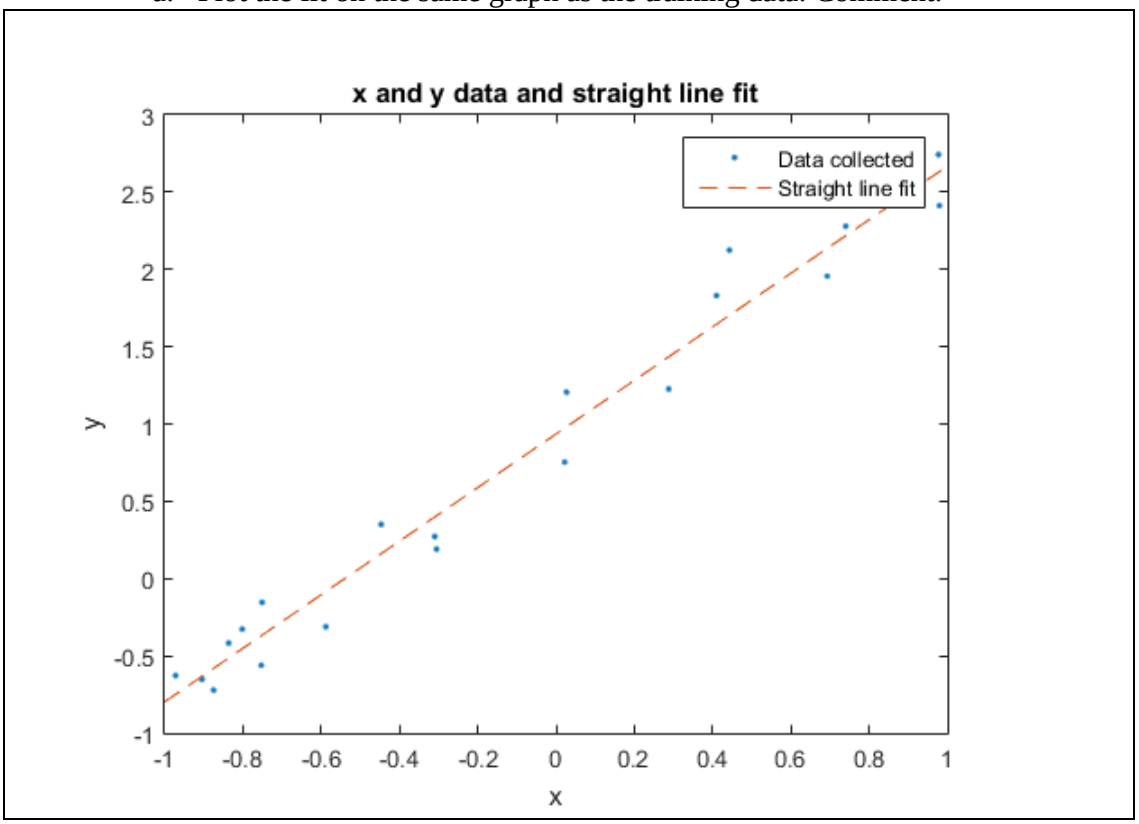
1. Write the matrix expressions for the LS estimate of the coefficients of a polynomial fit of degree P and of the corresponding sum of squares error, from training data $T=\{(x_i, y_i),\ i= 1,.., N\}$.



2. Write Matlab code to fit a polynomial of degree P to 1D data variables *x* and *y*. Write your own code, do not use any Matlab ready made function for LS estimation or for polynomial fitting. You should submit your code along with your report.

3. Load the data in file 'data1.mat' and use your code to fit a straight line to variables *y* and *x*.

    a. Plot the fit on the same graph as the training data. Comment.



Como se pode observar, há uma forte relação linear entre as variáveis *x* e *y*. Como tal, é esperado que a curva de regressão linear se ajuste de forma adequada, sendo por isso um modelo válido para a descrição dos pontos adquiridos

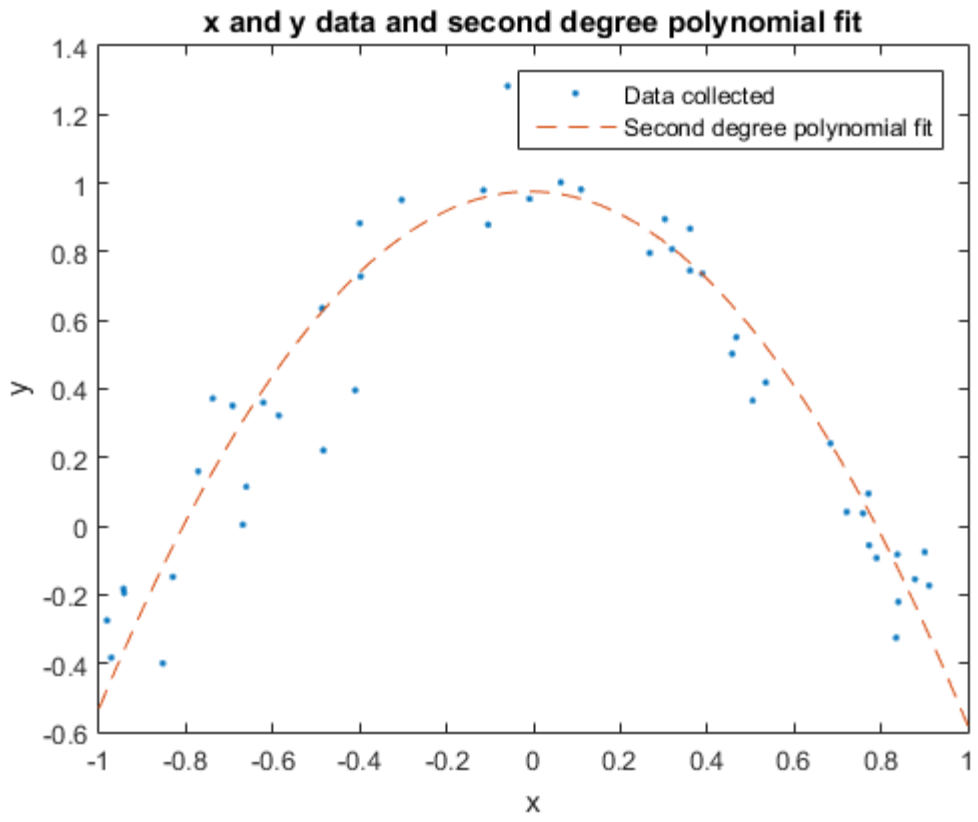
b. Indicate the coefficients and the error you obtained.

$\beta_1 = 1.7332$
$\beta_0 = 0.9351$
$\text{SSE}(\beta) = 0.7433$

Os valores dos coeficientes e do erro foram obtidos recorrendo à função desenvolvida na questão 1.2 (`polynomial_fit`), a qual segue em anexo ao relatório.

4. Load the data in file 'data2.mat', which contains noisy observations of a cosine function $= 2$ $+$, with $\in -1,1$, in which is Gaussian noise with a standard deviation of 0.15. Use your code to fit a second-degree polynomial to these data.

a. Plot the training data and the fit. Comment.



Tal como esperado, devido à série de Taylor do cosseno, este pode ser aproximado, em primeira hipótese, por uma função polinomial de grau 2 (na verdade, um polinómio cujo grau tende para infinito, no qual apenas são considerados os polinómios de grau par). Desta forma, dado as observações serem obtidas através de um cosseno (com ruído), seria de esperar que a curva de 2º grau se aproximasse dos pontos.
Tal é verificado através do plot obtido acima.

b. Indicate the coefficients and the error you obtained. Comment.

$\beta_2 = -1.5322$
$\beta_1 = -0.02571$
$\beta_0 = 0.9757$

$\text{SSE}(\beta) = 1.3415$

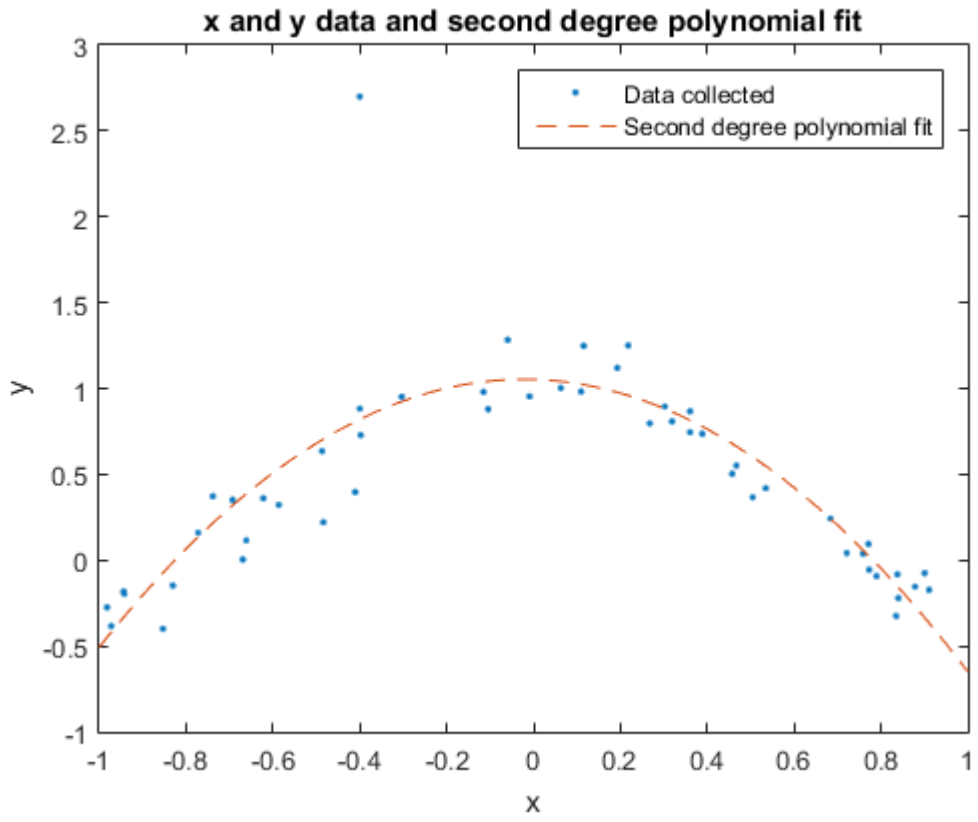
Os valores dos coeficientes e do erro foram obtidos recorrendo à função desenvolvida na questão 1.2 (`polynomial_fit`), a qual segue em anexo ao relatório.

Neste caso, o valor do erro (*sum of square error*) é bastante significativo. Tal é justificado devido ao menor ajuste que a curva de regressão aos pontos recolhidos. Dois dos motivos para um valor de erro tão elevado são a presença de ruído e a aproximação do cosseno a um polinómio de grau 2, o qual, como referido anteriormente, não é uma representação totalmente correta do mesmo.

5. Repeat item 4 using as input the data from file 'data2a.mat'. This file contains the same data used in the previous exercise except for the presence of an outlier point.

a. Plot the training data and the fit. Comment.

Este caso é em tudo semelhante ao anterior, excepto pela existência de um ponto mais "afastado" do restante conjunto. Dada a quantidade de pontos significativa, a existência de um único ponto que se afasta da tendência normal dos restantes, tal é "desprezado" pelo método dos *Least Squares*, sendo a curva obtida em tudo parecida à obtida anteriormente, como se pode verificar pelo plot e pelos valores dos coeficientes obtidos.



b. Indicate the coefficients and the error you obtained. Comment.

$\beta_2 = -1.63133$
$\beta_1 = -0.071597$
$\beta_0 = 1.05233$

$\text{SSE}(\beta) = 5.02487$

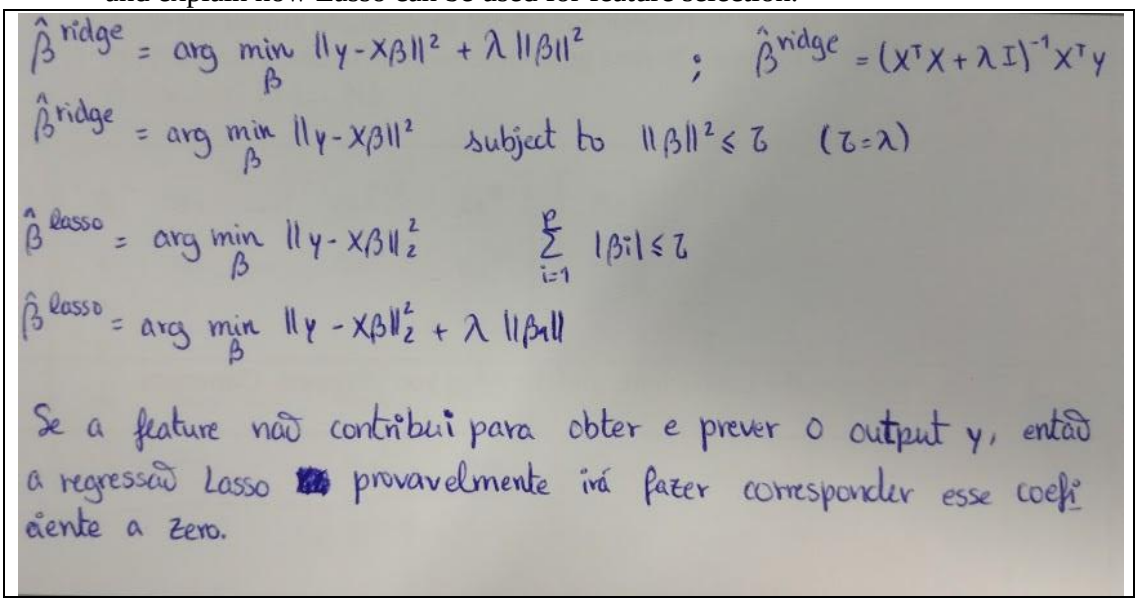
Os valores dos coeficientes e do erro foram obtidos recorrendo à função desenvolvida na questão 1.2 (`polynomial_fit`), a qual segue em anexo ao relatório.

Neste caso, o valor do erro (*sum of square error*) é superior ao anterior. Tal era esperado, pela maior distância de um dos pontos à curva obtida.

2. Regularization

The goal of this second part is to illustrate linear regression with regularization, we'll experiment with Ridge Regression and Lasso.

1. (T) Write the expression for the cost function used in Ridge Regression and Lasso and explain how Lasso can be used for feature selection.

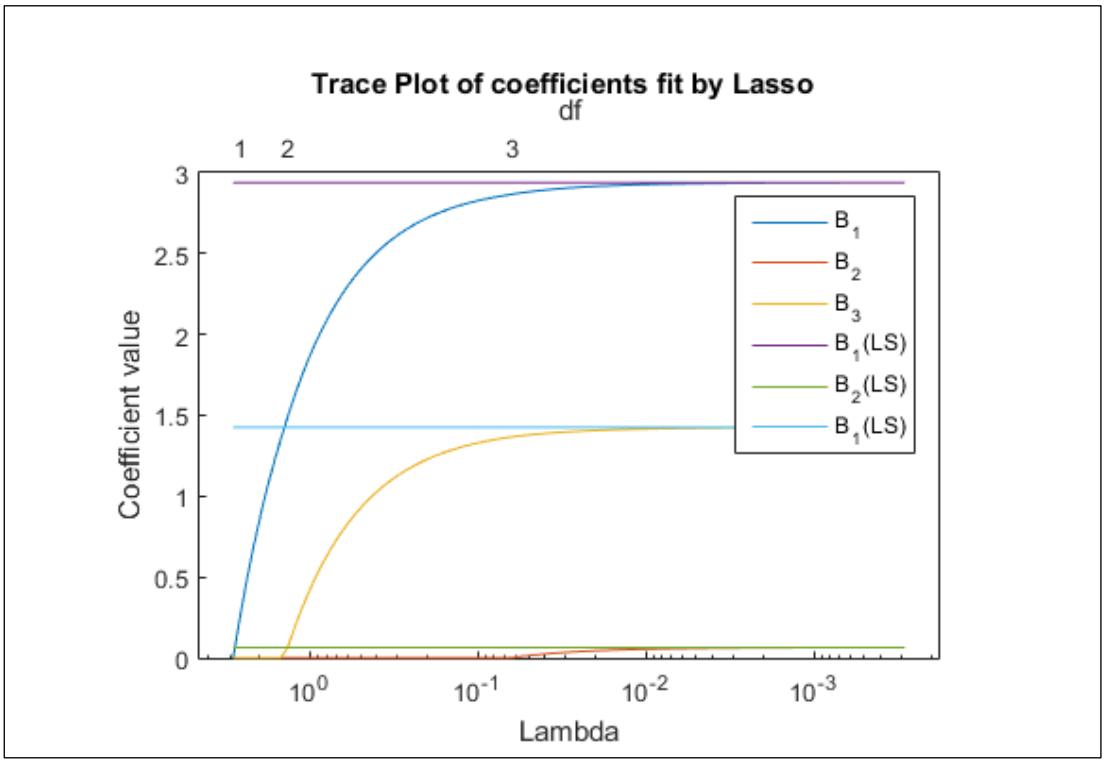


2. Load the data in file 'data3.mat' which contains 3-dimensional features in variable **x** and a single output *y*. One of the features in **x** is irrelevant. Use function `lasso` with default parameters (type help for more information on this function) and obtain regression parameters for different values of the regularization parameter λ (the values for lambda are returned in `FitInfo.Lambda`). Use function `lassoPlot` to plot the coefficients against λ. For comparison plot the LS coefficients in the same figure (λ = 0).

```
[B,FitInfo] = lasso(x,y);
lassoPlot(B,FitInfo,'PlotType','Lambda','XScale','log');
```

3. Comment on what you observe in the plot. Identify the irrelevant feature.

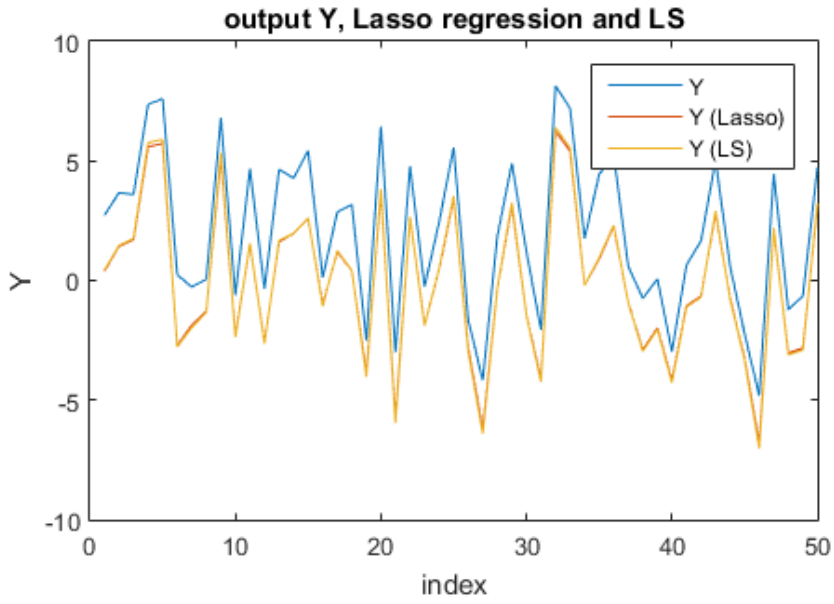
No gráfico abaixo observa-se a evolução dos coeficientes com λ. Nota-se que, à medida que λ tende para zero, os coeficientes aproximam-se dos mesmos obtidos pelo método dos mínimos quadrados.
Sabe-se que caso um dos parâmetros não contribua para o valor de y (seja irrelevante), a regressão de Lasso vai tentar aproximar o respetivo coeficiente de 0. Então, o parâmetro irrelevante corresponde a x_2.



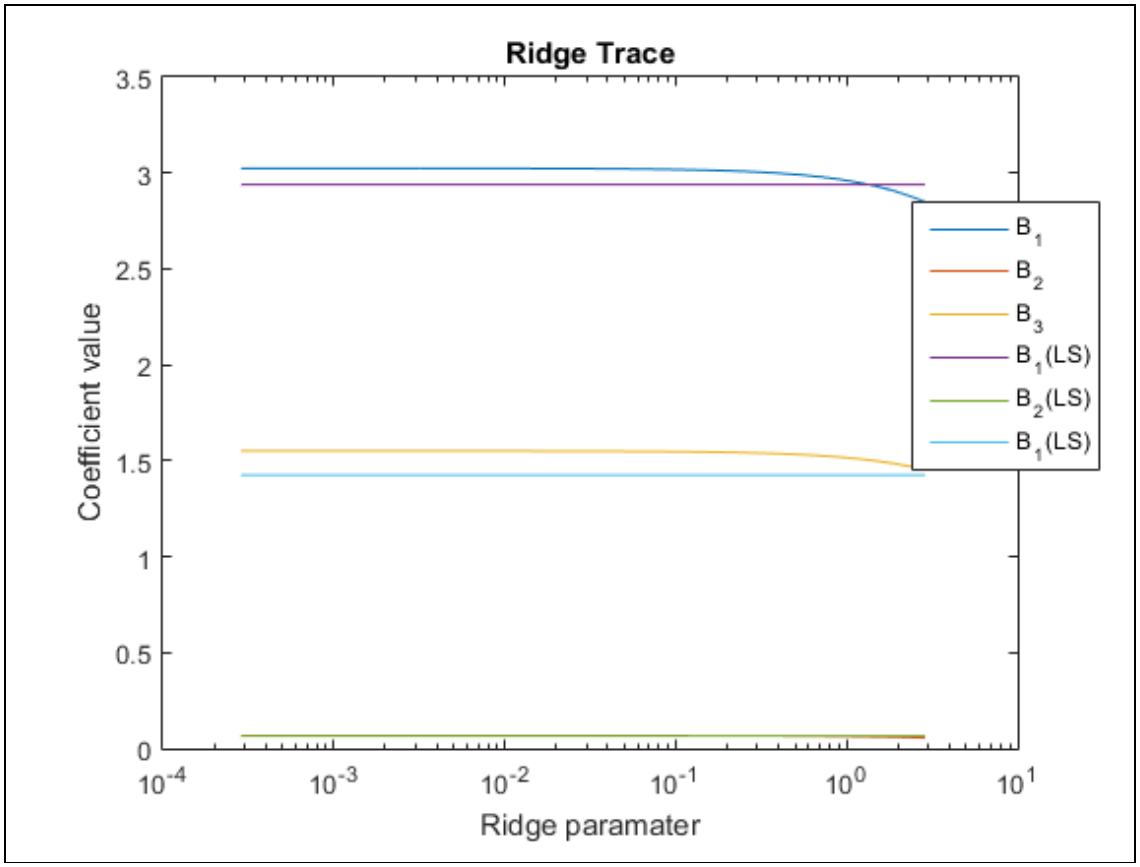
4. Choose an adequate value for λ. Plot *y* and the fit obtained for that value of λ. Compare with the LS fit. Compute the error in both cases. Comment.

Escolhe-se para λ o valor de 0.0633. Para este valor, observa-se que um dos parâmetros já tem o seu coeficiente nulo, estando os outros valores dos coeficientes ainda relativamente estáveis para pequenas variações de λ.
Nota-se que os resultados obtidos para os erros são bastante semelhantes, contudo há uma pequena diminuição do erro através da regressão de Lasso.
As estimativas de Y são bastante semelhantes, tal como é visível pelo gráfico abaixo, onde há uma grande sobreposição entre as estimativas de Y através do método dos mínimos quadrados Y (LS) e as estimativas de Y através do método de Lasso Y (Lasso).

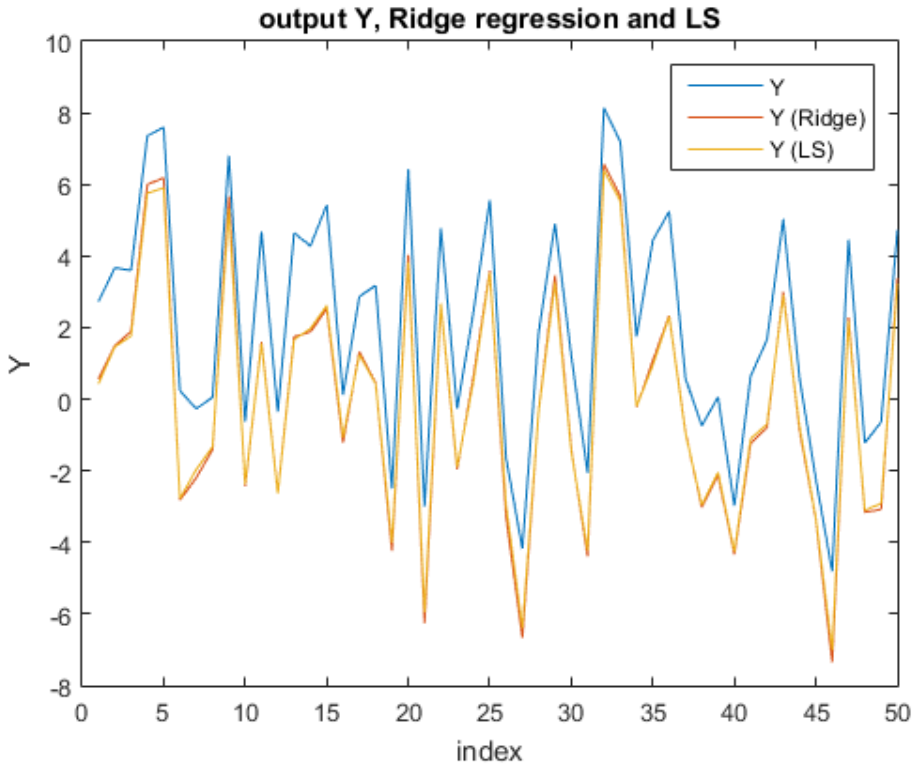
O valor de SSE para o método de LS é de 226,521 e para o método de Lasso é de 226,463.



5. Repeat the previous items but using ridge regression (function `ridge`) instead of Lasso. Use the same λ values as in Lasso.



O gráfico acima demonstra a evolução dos coeficientes com o parâmetro de Ridge. Estão também representados os valores dos coeficientes para o método LS.
Verifica-se também que não há nenhum valor do *Ridge parameter* que origine um coeficiente nulo. Contudo, pelo enunciado, continua-se a considerar um dos features irrelevantes.



SSE(Ridge)= 229,0398
SSE(LS)= 226,521

Verifica-se que, para este valor de λ, se obtém um SSE maior para o método de Ridge, pelo que a melhor escolha é Lasso.