Instituto Superior Técnico

Departamento de Engenharia Electrotécnica e de Computadores

Machine Learning

1st Lab Assignment

Shift: 5ª, 17h Group Number: 2

Number 81567 Name José Pedro Mincgoja

Number 81138 Name João Remino

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

$$f(\mathbf{x}) = \beta_0 + \sum_{k=1}^{P} \beta_k x_k$$

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

$$SSE = \sum_{i=1}^{N} (y_i - f(\mathbf{x}^i))^2$$

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, $\phi_k(\mathbf{x})$, of the input variables

$$f(\mathbf{x}) = \beta_0 + \sum_{k=1}^{K} \beta_k \phi_k(\mathbf{x})$$

In this case the model is still linear in the parameters although it is nonlinear in $\mathbf{x}$. 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 $T=\{(x_i, y_i), i= 1,\ldots, N\}$.

$$X = \begin{bmatrix} 1 & x_1 & \cdots & (x_1)^P \\ 1 & x_2 & \cdots & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & x_{N-1} & & \\ 1 & x_N & & (x_N)^P \end{bmatrix} \qquad [\hat{\beta}] = (X^T X)^{-1} X^T y$$

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.

(ver plot em anexo)
Como se pode observar, a estimativa resulta numa reta (pois degree P=1), cuja aproximação é fiável pois minimiza a Sum of Squared Errors.

   b. Indicate the coefficients and the error you obtained.

$[\beta] = \begin{bmatrix} 0,9351 \\ -1,7332 \end{bmatrix}$ erro: 0,7433

4. Load the data in file 'data2.mat', which contains noisy observations of a cosine function $y = cos(2x) + \varepsilon$, with $x \in [-1,1]$, in which ε 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.

(ver plot em anexo)
Como se pode ver, devido ao grau ser 2, a aproximação obtida é parabólica e esta ~~que se~~ consegue se adaptar a função y devido a esta ser par (Cos) e quase com forma parabólica no intervalo usado.

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

$[\beta] = \begin{bmatrix} 0,9757 \\ -0,0257 \\ -1,5322 \end{bmatrix}$ erro: 1,3416 Devido ao coseno no intervalo considerado, não ser efetivamente uma parábola, o erro associado será maior.

$\beta[0]$ é próximo de 1 pois a parábola na origem é 1 tal como cos(0)
$\beta[1]$ é próximo de 0 pois o coseno não tem componentes ímpares na sua série de McLaurin pois é uma função par.
$\beta[2]$ é negativo a parábola está virada para baixo.

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.

Pode-se observar que a função se encontra mais acima que a anterior por influência do ponto outlier. De resto os parâmetros β são praticamente iguais.

(ver plot em anexo)

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

$$\beta = \begin{bmatrix} 1{,}0523 \\ -0{,}0716 \\ -1{,}6313 \end{bmatrix}$$

erro: 5,0249

Devido ao outlier o erro obtido será muito superior.
Os coeficientes são bastante semelhantes.
O $\beta(0)$ aumentou pois o outlier atraiu a função para cima.

2. Regularization

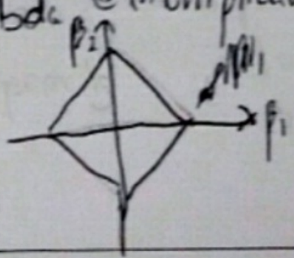
The goal of this second part is to illustrate linear regression with regularization, we'll experiment with Ridge Regression and Lasso.

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.

Ridge: $\|y - X\beta\|^2 + \lambda\|\beta\|^2$

Lasso: $\|y - X\beta\|_2^2 + \lambda\|\beta\|_1$

A regressão de Lasso é usada para "feature selection" pois o lambda é multiplicado por $\|\beta\|_1$ ou seja $\|\beta\|_1 \le \tau$



O que gera um vetor esparso de coeficientes.

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 ($\lambda = 0$).

```
[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.

Pode-se observar, a evolução dos coeficientes β à medida que lambda diminui.

A feature irrelevante será o β_2 pois com a variação de λ este toma valores bastante inferiores aos outros coeficientes chegando a ser zero.

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.

O valor de λ adequado é 0,1762 pois vai eliminar 1 das features. (este valor foi obtido através de cross validation).

erro Lasso: 18,55

erro LS fit: 14,98

Neste caso pode se entender o LS fit aproxima-se melhor aos pontos que o Lasso.

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

erro ridge : 16,0212
erro LS fit : 14,98

Daqui se pode entender que neste ~~[illegible]~~ caso a ridge regression aproxima melhor aos pontos, do que o lasso.