MCIT 515

Fundamentals of Linear Algebra and Optimization
Jean Gallier and Jocelyn Quaintance

Project 5: Ridge Regression

The purpose of this project is to implement several versions of ridge regression. You will not need to write a report for this project, the functions will be tested by the auto-grader and the images will be generated from the output script like in the first few projects.

Recall that ridge regression for learning an affine function $f(x) = x^\top w + b$ from the training data $((x_1, y_1), \ldots, (x_m, y_m))$ is the following optimization problem:

Program (RR3):

$$\begin{aligned} &\text{minimize} \quad \xi^\top \xi + K w^\top w \\ &\text{subject to} \\ &\qquad\qquad y - Xw - b\mathbf{1}_m = \xi, \end{aligned}$$

with $y, \xi, \mathbf{1}_m \in \mathbb{R}^m$ and $w \in \mathbb{R}^n$ ($\mathbf{1}_m$ is the vector (of dim m) whose components are all equal to 1) and $K > 0$ a fixed constant.

Here X is an $m \times n$ matrix whose rows are the transpose of the data points $x_1, \ldots, x_m \in \mathbb{R}^n$, and $y \in \mathbb{R}^m$.

The first solution is obtained by centering the data: the centered data are $\widehat{y} = y - \overline{y}\mathbf{1}_m$ and $\widehat{X} = X - \overline{X}$, where $\overline{X}$ is the $m \times n$ matrix whose jth column is $\overline{X^j}\mathbf{1}_m$, the vector whose coordinates are all equal to the mean $\overline{X^j}$ of the jth column X^j of X.

The optimal solution w is given by

$$w = \widehat{X}^\top (\widehat{X}\widehat{X}^\top + K I_m)^{-1} \widehat{y}, \qquad (*_{w_6})$$

and b is given by

$$b = \overline{y} - (\overline{X^1} \ \cdots \ \overline{X^n}) w,$$

where $(\overline{X^1} \ \cdots \ \overline{X^n})$ is the $1 \times n$ row vector consisting the the means of the columns of X.

(Part 1) (**20 points**) Write a `Matlab` function `ridgeregv1` to compute w and b from X and y. This function takes X, y, and $K > 0$ as input, and returns w, the Euclidean norm nw of w, b, the error vector $xi = \widehat{y} - \widehat{X}w$, and the Euclidean norm nxi of xi.

```
function [w,nw,b,xi,nxi] = ridgeregv1(X,y,K)
%  Ridge regression with centered data
%  b is not penalized
```

```
%  X is an m x n matrix, y a m x 1 column vector
%  weight vector w, intercept b
%  Solution in terms of the primal variables
%
m = size(y,1);
n = size(X,2);
%
% Your code
%
end
```

(Part 2) (**20 points**) The dual of Program (**RR3**) is

Program (**DRR3**):

$$\begin{aligned} &\text{minimize} \quad \alpha^\top (XX^\top + KI_m)\alpha - 2\alpha^\top y \\ &\text{subject to} \\ &\qquad\qquad \mathbf{1}^\top \alpha = 0, \end{aligned}$$

where the minimization is over α. This program can be solved directly without centering the data by solving the KKT equations

$$\begin{pmatrix} XX^\top + KI_m & \mathbf{1}_m \\ \mathbf{1}_m^\top & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \mu \end{pmatrix} = \begin{pmatrix} y \\ 0 \end{pmatrix}.$$

Then we have

$$\begin{aligned} w &= X^\top \alpha \\ b &= \mu \\ \xi &= K\alpha. \end{aligned}$$

Write a `Matlab` function `ridgeregb1` to compute w, b, α and ξ from X and y. This function takes X, y, and $K > 0$ as input, and returns w, b, the error vector xi, the Euclidean norm nxi of xi, and α.

```
function [w,b,xi,nxi,alpha] = ridgeregb1(X,y,K)
%  Ridge regression
%  b is not penalized
%  X is an m x n matrix, y a m x 1 column vector
%  weight vector w, intercept b
%  Solution in terms of the KKT equations
%
m = size(y,1);
n = size(X,2);
```

```
X1 = X*X' + K*eye(m);
%
% Your code
%
end
```

Compare the solutions for w and b given by `ridgerev1` and `ridgeregb1` (they should agree up to roughly ten decimals).

(Part 3) (**20 points**) Another way to solve ridge regression is to penalize b. This corresponds to the following optimization problem:

$$\begin{aligned}
&\text{minimize} \quad \xi^\top \xi + K w^\top w + K b^2 \\
&\text{subject to} \\
&\qquad\qquad y - Xw - b\mathbf{1}_m = \xi,
\end{aligned}$$

minimizing over ξ, w and b.

This suggests treating b as an extra component of the weight vector w and by forming the $m \times (n+1)$ matrix $[X\ \mathbf{1}]$ obtained by adding a column of 1's (of dimension m) to the matrix X, and we obtain

Program (RR3b):

$$\begin{aligned}
&\text{minimize} \quad \xi^\top \xi + K w^\top w + K b^2 \\
&\text{subject to} \\
&\qquad\qquad y - [X\ \mathbf{1}] \begin{pmatrix} w \\ b \end{pmatrix} = \xi,
\end{aligned}$$

minimizing over ξ, w and b.

It can be shown that the solution is given by

$$\begin{aligned}
\alpha &= ([X\ \mathbf{1}][X\ \mathbf{1}]^\top + K I_m)^{-1} y \\
\begin{pmatrix} w \\ b \end{pmatrix} &= [X\ \mathbf{1}]^\top \alpha \\
\xi &= K\alpha.
\end{aligned}$$

Thus $b = \mathbf{1}^\top \alpha$.

Write a `Matlab` function `ridgeregv2` to compute w and b from X and y. This function takes X, y, and $K > 0$ as input, and returns w, the Euclidean norm nw of w, b, the error vector $xi = K\alpha$, and the Euclidean norm nxi of xi.

```
function [w,nw,b,xi,nxi] = ridgeregv2(X,y,K)
%  Ridge regression minimizing w and b
%  b is penalized
%  X is an m x n matrix, y a m x 1 column vector
%  weight vector w, intercept b
%  Solution in terms of the dual variable
%
m = size(y,1); n = size(X,2);
XX = [X ones(m,1)];
%
% Your code
%
end
```

(Part 4) (**20 points**) As a least squares problem, the solution is given in terms of the pseudo-inverse $[X\ \mathbf{1}]^{+}$ of $[X\ \mathbf{1}]$ by

$$\begin{pmatrix} w \\ b \end{pmatrix} = [X\ \mathbf{1}]^{+} y.$$

Write a `Matlab` function `reglq` to compute w and b from X and y. This function takes X, y as input, and returns w, the Euclidean norm nw of w, b, the error vector $xi = y - Xw - b\mathbf{1}$, and the Euclidean norm nxi of xi.

```
function [w,nw,b,xi,nxi] = reglq(X,y)
%  Regression minimizing w and b
%  X is an m x n matrix, y a m x 1 column vector
%  weight vector w, intercept b
%  Computes the least squares solution using the pseudo inverse
%  Use pin
%
m = size(y,1); n = size(X,2);
XX = [X ones(m,1)];
%
% Your code
%
end
```

(Part 5) (**20 points**) To test your four functions we will also be looking at images based on the output of your scripts. These will constitute the rest of your grade.