Learning outcomes

1. To revise basic matrix manipulations by hand.
2. To learn how to carry out basic matrix manipulations in MATLAB.
3. To learn how to plot simple graphs.
4. To pull together basic skills to implement ridge regression.

Basic matrix calculations

Given the matrices $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \\ 3 & 2 \end{pmatrix}$ and $w = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$,

1) What is Aw?

2) What are the dimensions of the resulting matrix?

3) What is wA?

4) What is A^T (aka A')?

5) What is the identity matrix

6) What is $A \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$?

7) Rearrange the following equation $w_1x + w_2y = 5$ into the format $y = mx + c$.

Ridge regression exercise

In this exercise you will implementing the primal ridge regression algorithm and apply it to a small data set. The primal ridge regression formula is:

$$w_{opt} = (Z'Z + CI)^{-1}Z'y \tag{1}$$

C is a constant called the ridge coefficient which determines how much slack we allow in the solution for the optimal weights w_{opt}. These weights determine the best fit line for the training data points.
To find the corresponding prediction of label $\hat{y}$, for a new test example x, we just calculate $(w_{opt} \cdot x) = w'x$.
Commonly used notation for the the training set is as follows:
We think of the training set as a sequence of l examples, represented as information pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)$. Each example contains an n dimensional feature vector $x_i = \begin{pmatrix} x_i^1 \\ x_i^2 \\ \vdots \\ x_i^n \end{pmatrix}$ (which describes each example), with a corresponding label y_i (which is the target to be learnt). With ridge regression problems y is a continous real value.
The matrix Z mentioned in the ridge regression formula (1) above is of the form:

$$\begin{pmatrix} 1 & x_1^1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2^1 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_l^1 & x_l^2 & \cdots & x_l^n \end{pmatrix} \tag{2}$$

The matrix Z is an $l \times (n+1)$ matrix, which defines the training set feature vectors as its rows, augmented with an extra column of 1's to add an extra constant term w_0 (sometimes known as the bias) to the solution found by the ridge regression algorithm. This means that the resultant solution of optimal weights is an $(n+1) \times 1$ vector $w_{opt} = \begin{pmatrix} w^0 \\ w^1 \\ w^2 \\ \vdots \\ w^n \end{pmatrix}$.

1) Show how the ridge regression formula gives a ridge regression formula (1) gives a $(n+1) \times 1$ vector solution.

You will now write a program in MATLAB that will perform ridge regression on a simple data set. This program will take several lines of code so create a '.m' file to store it in by clicking on **File→New→M-file**. To run the program at any time click on **Debug→Run**, and view the output in the command window. For each part write down the corresponding MATLAB code in the gap provided, with a brief description of what it is doing. Remember if you want to find out more about a MATLAB routine, then type
`help` *routinename* in the command window.

2) Create the training data matrices $x = \begin{pmatrix} 3 \\ 1 \\ 9 \\ 7.2 \\ 10.4 \\ 5 \end{pmatrix}$ which contains the features of the examples in its rows, $y = \begin{pmatrix} 5 \\ 3.5 \\ 13 \\ 11 \\ 18 \\ 6 \end{pmatrix}$. What is the number of features n and the number of examples l in this data?

3) Create the matrix Z from adding an extra column of ones to the feature matrix x. Check the dimensions of these matrices using MATLAB. (Hint: use the MATLAB routine `ones`, and concatenate the matrices to create Z, then use `size` to check the dimensions).

4) Plot these data points on a graph as a set of blue crosses. (Hint: use the `plot` routine, with matrices x and y).

5) Create a 2×2 identity matrix called I. (Hint: use the `eye` routine).

6) Calculate Z'. What are its dimensions?

7) Create a variable C and set it equal to 1. Now using the previous parts calculate the inside brackets of the ridge regression equation $(Z'Z + CI)$ and store it in a matrix called $BRACKETS$.

8) Calculate the inverse of the matrix $BRACKETS$, and store it in the matrix called $INVBRACKETS$.

9) Using the $INVBRACKETS$, Z' and y matrices, and the ridge regression formula (1) to calculate the optimal weights and store them in a matrix w.

10) Calculate label $\hat{y}$ for the new test point which has the feature x=6, using the optimal weights w. (Hint: use the formula detailed earlier, to calculate the label $\hat{y}$ for a new point).

11) Find the minimum and maximum of the feature x values, and store them in variables named $minx$ and $maxx$ accordingly. (Hint: use the routines `min` and `max`).

12) Create a matrix $RANGE$, which contains a set of values which start from $minx$ to $maxx$ in increments of 0.2. (Hint: use the colon : notation to build the range).

13) Add an extra row of 1's to the matrix $RANGE$. (Hint: You may need to find the `size` of matrix range and then use the routine `ones`).

14) Using the matrix $RANGE$ and the optimal weights w calculated earlier, calculate a matrix f that will hold the best fit line coordinates. (Hint: use the same technique as in Question 10).

15) Plot this $RANGE$ (taking out the row of 1's) against the function values f, to see the best fit line. Modify the code so that you can superimpose this line over the data points plotted earlier. (Hint: use the `hold on` and `hold off` commands).

16) Try to change the ridge coefficient C to see how the resulting best fit line changes. Add titles and suitable axis titles to the graph. (Hint: use the commands `title`, `xlabel` and `ylabel`).

17) Try to think of some examples of real life problems that can be formulated as ridge regression problems.

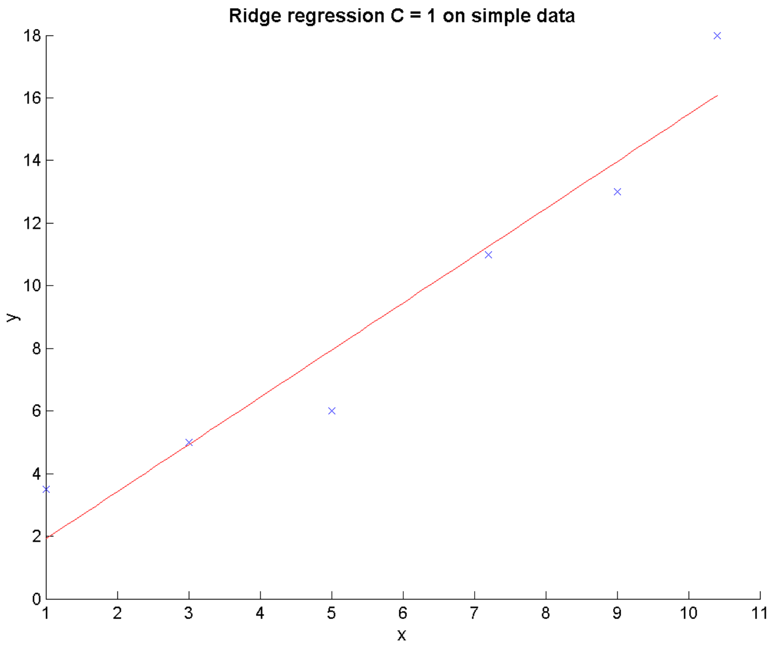

Figure 1 The finished ridge regression plot: After completing this exercise you should finish with a plot that looks like this.