COMPGV17 GLM Practical 1

Gary Hui Zhang, PhD

Task 1: Variable Space vs Subject Space

Three individuals take a two-part exams. The first person received a score of 5 and 3 on the two parts, the second scored a 7 and 2, and the third a 4 and 6.

1. Represent these results as points in variable space and then as points in subject space. Create the respective plots with Matlab. For each space, write down its dimension (i.e., the number of axes) and the number of points in the space; explain what each dimension (axis) represents.

2. Compute the mean score for each part of the exam averaged over the three individuals and centre the scores from each part by subtracting the respective mean.

3. Repeat 1.1 with the centred scores.

4. For each point in the subject space, compute the length of the centered vector pointing from the origin to the point. Compare this length to the standard deviation of the corresponding scores. What relationship do you find?

5. For the points in the subject space, compute the cosine of the angle between the corresponding vectors. Compare this value to the correlation coefficient between the scores from two parts. What do you find?

Task 2: Simple Bivariate Regression Example

Consider the following small set of bivariate data:

X	10	5	1	6	7	3	4	5	1	8
Y	2	4	4	2	4	5	4	5	6	4

1. Compute the means, standard deviations, and correlation between the variables and plot them as a scatterplot.

2. Center the scores for both variables and repeat 2.1.

3. Now consider the two variables, after centering by the means, in the subject space. What is the dimension of the space? Compute the length of the corresponding vectors and the cosine of the angle between the vectors. Relate these to the answer to 2.1.

4. Assume the variable X as the predictor and Y as the response variable, estimate the regression coefficient, the predicted response, and the error between the measured and the predicted response.

5. Plot the line predicted by the estimated regression coefficient overlaid over the scatterplot for the centred variables. Mark out the points along the line corresponding to the predicted values of Y for the given set of X values. From this, deduce how the errors are represented in the plot.

6. Compute the ratio of the lengths of Y and X. How does the correlation coefficient between X and Y relate to the regression coefficient estimated in 2.4 and this ratio?

7. Reversing the role of the two variables, repeat 2.4. What can you learn?