# Advanced Regression: Multiple choice

25th April 2019

## Sums

The sum operator is defined as

$$\sum_{i=1}^{n} i = 1 + 2 + 3 + .... + (n-1) + n \tag{1}$$

Often it is used to iterate through elements of vectors or matrices

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + .... + x_{(n-1)} + x_n \tag{2}$$

In the advanced regression classes you will commonly see two sorts of sums:

1. Sum over samples $i = 1, ..., n$
   For example in regression we can compute the fitted values $\hat{y}_i$ for each observation $i$, i.e. the values of the outcome predicted for observation $i$ based on the linear model using data $x_i$. In matrix notation the fitted values for all observations $\hat{y} = (\hat{y}_1, ...., \hat{y}_n)$ can be computed with the hat matrix $h$ of dimension $n \times n$

   $$\hat{y} = x\hat{\beta} = \underbrace{x}_{n \times p} \underbrace{(x^t x)^{-1}}_{p \times p} \underbrace{x^t}_{p \times n} y = \underbrace{h}_{n \times n} y.$$

   Then the residual sums of squares are defined as

   $$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (e_i)^2,$$

   where $e_i = (y_i - \hat{y}_i)$.

2. Sum over variables $j = 1, ..., p$
   For example the penalty term of regularised regression is a sum of regresssion coefficient estimates, Ridge for example uses the following L2 penalty

   $$\lambda f(\beta) = \lambda \sum_{j=1}^{p} \beta_j^2.$$

## Matrices and matrix multiplication

In the following we consider $x$ to be a data matrix of dimension $n \times p$, with $n$ observations and $p$ variables. When performing matrix multiplication it is extremely important to keep track of the dimension of the matrices and only multiply matrices with the correct dimension. Note how the dimensions in the least squares estimate $\hat{\beta}_{OLS}$ are defined as

$$\hat{\beta}_{OLS} = \underbrace{(x^t x)^{(-1)}}_{p \times p} \underbrace{x^t}_{p \times n} \underbrace{y}_{n \times 1} \tag{3}$$

and produce an estimate that is of dimension $p \times 1$, one regression coefficient per variable.

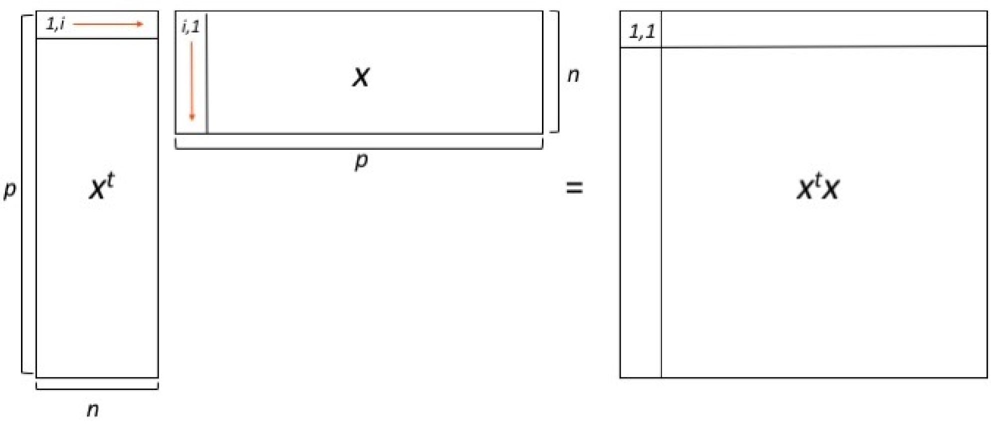
Importantly, for matrices the order matters, for example the between-variables covariance matrix (dimension $p \times p$) is defined as

$$cov(x^t) = \frac{1}{n} x^t x, \tag{4}$$

while the between-observations covariance matrix (dimension $n \times n$) is defined as

$$cov(x) = \frac{1}{p} x x^t. \tag{5}$$

To compute the element in the first row and first column of the matrix $x^t x$ (proportional to the $p \times p$ variable-covariance matrix), we take the first row of $x^t$ and the first column of $x$ as shown in the figure below.



Then $x^t x[1,1]$, the element in the first row and first column of the matrix $x^t x$, is computed as

$$x^t x[1,1] = \sum_{i=1}^{n} x^t_{1,i} x_{i,1} \tag{6}$$

where we sum over individuals, the sum goes from $i = 1, ..., n$.

Equivalently, $xx^t[1,1]$, the first element in $xx^t$ (proportional to the $n \times n$ observation-covariance matrix) first row and first column of the matrix $xx^t$, is computed as

$$xx^t[1,1] = \sum_{j=1}^{p} x_{1,j} x^t_{j,1} \tag{7}$$

where we sum over variables, the sum goes from $j = 1, ..., p$.

More intuition on the matrix multiplication is given here `https://www.math3ma.com/blog/matrices-probability-graphs`.