Some notes on our matrix notation

We will introduce some notation here. First consider a random vector $\mathbf{y} = (y_0, \ldots, y_n)^T = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix}$. Note that in some cases **y** is considered a fixed vector, but here we consider it to be random, i.e. that the entries y_i are random variables. In our notation we will always denote the mean of a quantity by $E(a) = \overline{a}$. Thus for the random vector **y**, we write

$$\overline{\mathbf{y}} = E(\mathbf{y}) = (\overline{y}_0, \ldots, \overline{y}_n).$$

We also denote the variance of a single random variable X to be $V(y) = E[(y - \overline{y})^2]$.

The same notation, however, is also used for the variance (more properly the *covariance matrix*) of the random vector $\mathbf{y} = (y_0, \ldots, y_n)$.

For now assume that the mean $\overline{\mathbf{y}} = 0$; if this is not the case just replace **y** in any formula by $\mathbf{y} - \overline{\mathbf{y}}$.

We define the *variance* or *covariance matrix* of the random vector **y** to be the matrix $V(\mathbf{X})$ whose i, j entry is (recall we are assuming $\overline{y}_i = 0$)

$$V(\mathbf{y})_{ij} = \mathrm{Cov}(y_i, y_j) = E\big[y_i - \overline{y}_i)(y_j - \overline{y}_j)\big] = E(y_i y_j).$$

Now consider the matrix

$$\mathbf{y}\mathbf{y}^T = (y_0, y_1, \ldots, y_n) \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} y_0^2 & y_0 y_1 & \cdots & y_0 y_N \\ y_1 y_0 & y_1^2 & \cdots & y_1 y_N \\ \vdots & \vdots & \ddots & \vdots \\ y_N y_0 & y_N y_1 & \cdots & y_N^2 \end{bmatrix}.$$

We have

$$E(\mathbf{y}\mathbf{y}^T) = \begin{bmatrix} E(y_0)^2 & E(y_0 y_1) & \ldots & E(y_0 y_N) \\ E(y_1 y_0) & E(y_1^2) & \ldots & E(y_1 y_N) \\ \vdots & \vdots & \ddots & \vdots \\ E(y_N y_0) & E(y_N y_1) & \ldots & E(y_N^2) \end{bmatrix}$$

Therefore,

$$E(\mathbf{y}\mathbf{y}^T)_{ij} = E(y_i y_j) = V(\mathbf{y})_{ij},$$

so that we can also write

$$V(\mathbf{y}) = E(\mathbf{y}\mathbf{y}^T).$$

A basic identity:

Note also that if **A** is a matrix, then by above

$$V(\mathbf{A}\mathbf{y}) = E\left[(\mathbf{A}\mathbf{y})(\mathbf{A}\mathbf{y})^T\right] = E(\mathbf{A}\mathbf{y}\mathbf{y}^T\mathbf{A}^T) = \mathbf{A}E(\mathbf{y}\mathbf{y}^T)\mathbf{A}^T = \mathbf{A}V(\mathbf{y})\mathbf{A}^T.$$

Since $V(\mathbf{y}) = V(\mathbf{y} - \overline{\mathbf{y}})$ for any random vector **y**, the above identity

$$V(\mathbf{A}\mathbf{y}) = \mathbf{A}V(\mathbf{y})\mathbf{A}^T$$

also holds for random vectors **y** with non-zero mean.