Basic Linear Algebra Operations and Definitions

Here I want to introduce some notation that I use (which may be non-standard) and some ideas which are basic and well known, but helpful to keep in mind when reading the ML literature. The vector $\mathbf{u}$ with n elements is either a horizontal or vertical collection of n numbers. In the former $\mathbf{u}$ is *row* vector (and I note its dimension $\mathbf{u} \equiv (1 \times n)$) and the latter it is a *column* vector ($\mathbf{u} \equiv (n \times 1)$). By default all vectors are column vectors. A column vector $\mathbf{u}$ can be written as a row by tranposing (ie $\mathbf{u}^T$). We define the two vector-vector products:

Vector Products

The *dot product* of two vectors $\mathbf{u}$ and $\mathbf{v}$ is the scalar

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} = \sum \mathbf{u}_i \mathbf{v}_i \tag{1}$$

The *vector product* of two vectors $\mathbf{u}$ and $\mathbf{v}$ is the matrix

$$\mathbf{u}\mathbf{v}^T = \begin{pmatrix} u_1 v_1 & u_1 v_2 & \ldots & u_1 v_n \\ u_2 v_1 & u_2 v_2 & \ldots & u_2 v_n \\ \vdots & \vdots & \ddots & \vdots \\ u_n v_n & u_1 v_n & s & u_n v_n \end{pmatrix} \tag{2}$$

The vectors $\mathbf{u}$ A $m \times n$ matrix A, where

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \tag{3}$$

is often decomposed into its m rows

$$A = \begin{pmatrix} \mathtt{r}(A)_1 \\ \mathtt{r}(A)_2 \\ \vdots \\ \mathtt{r}(A)_m \end{pmatrix} \tag{4}$$

or n columns

$$A = \begin{pmatrix} \mathtt{c}(A)_1 & \mathtt{c}(A)_2 & \ldots & \mathtt{c}(A)_n \end{pmatrix} \tag{5}$$

so as to better represent matrix operations. With this notation we view the matrix product AB where $A \equiv (m \times n)$ and $B \equiv (n \times m)$ as a matrix of dot products

$$AB = \begin{pmatrix} \mathtt{r}(A)_1 \cdot \mathtt{c}(B)_1 & \mathtt{r}(A)_1 \cdot \mathtt{c}(B)_2 & \cdots & \mathtt{r}(A)_1 \cdot \mathtt{c}(B)_m \\ \mathtt{r}(A)_2 \cdot \mathtt{c}(B)_1 & \mathtt{r}(A)_2 \cdot \mathtt{c}(B)_2 & \cdots & \mathtt{r}(A)_2 \cdot \mathtt{c}(B)_m \\ \vdots & \vdots & \ddots & \vdots \\ \mathtt{r}(A)_m \cdot \mathtt{c}(B)_m & \mathtt{r}(A)_1 \cdot \mathtt{c}(B)_m & \cdots & \mathtt{r}(A)_m \cdot \mathtt{c}(B)_m \end{pmatrix} \tag{6}$$

The matrix-vector product $A\mathbf{x}$ is the weighted sum of the columns of A ie

$$A\mathbf{x} = \sum x_i \mathtt{c}(A)_i \tag{7}$$

The set of all possible combinations of the columns defines the *column space* of A. This idea is couple to the idea of a *vector space*. More specifically, the *vector space* generated by a set of n vectors

$\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is the set of all possible linear combinations of the $\mathbf{u}_i$. If we combine these vectors to form the columns of a matrix A, the we get the linear combination from the above matrix-vector product $A\mathbf{x}$. The column space of a matrix A then is the *vector space* generated by its columns. This *vector space*, which is the *column space* of A is often also called the *range* of A, ie

$$\text{range}\,(A) = \{\text{all possible} A\mu x_i\} \tag{8}$$

since it defines which vectors can be produced by summing the columns in A. The *null space* is a particular set of non-zero weights $\mathbf{x}$ which produce the zero vector, ie

$$\text{null}\,(A) = \{\text{all}\mathbf{x}_i\text{where}A\mathbf{x}_i = 0.\} \tag{9}$$

If such an $\mathbf{x}_i$ exists, then one column can be written as a linear combination of at least one of the others. This has significant implications - see below. If no such combination exists, then the columns are *linearly independant* - one cannot be written as a combination of the others. This concept is again common to sets of vectors - ie a set of vectors is linearly independant if no linear combination of the vectors produces a zero vector. Spaces which are *not* linearly independant include redundant vectors.

Consider a matrix A and some set of vectors $\begin{pmatrix} u_1 & u_2 & \ldots & u_n \end{pmatrix}$. For every possible combination of the columns of A (ie every $v = A\mathbf{x}$), it may be possible to choose some combination of the vectors $\mathbf{u}_i$ that produces the same v, and vica versa. In this sense, the set of vectors$\begin{pmatrix} u_1 & u_2 & \ldots & u_n \end{pmatrix}$ can substitute for the matrix A, since they can both generate the exact same output (ie they have the same range). If we want to do this substitution, we shouldnt choose any redundant vectors in the $\mathbf{u}_i$ - we should only use *linearly independant* vectors. Such a set of vectors is said to form a *basis* for A.