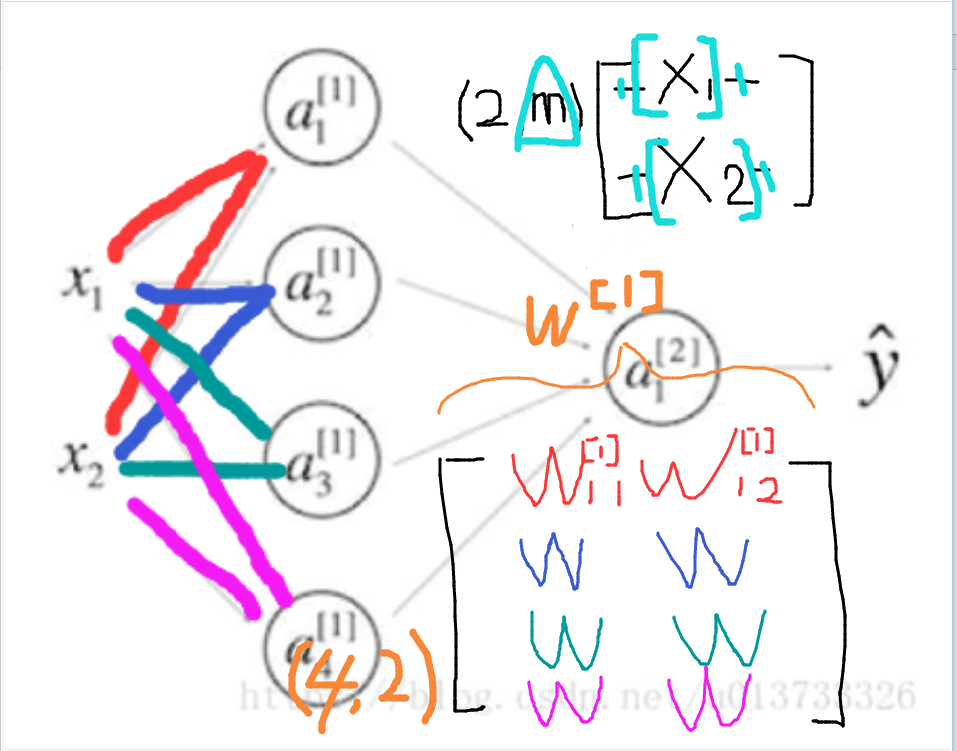
对于w矩阵，行数：本层神经元的数量，列数：前一层神经元的数量(第0层即是输入层特征的数量)

对于b矩阵，行数：该层w矩阵的行数，列数：1

对于z和a矩阵，行数：该层w矩阵的行数，列数：1



**n【0】=2 n【1】=4 n【3】=1**

X/A【0】： **（n【0】，m） （2，m）**

W­【1】 = 输入层到第一层的参数个数 **（n【1】，n【0】） (4，2）**

每一根线就代表一个参数

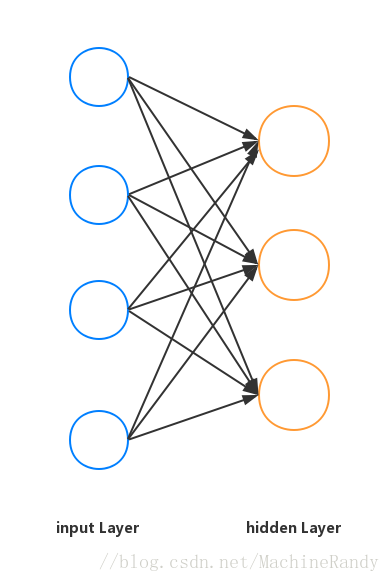
b【1】 =W的个数 **（n【1】，0） (4，1)**

**A[1]**  =第一层的层数  **（n【2】，1） （4，1）**

**A[1] =σ（WX+b）**

**=σ（(4,2)(2,m)+(4,1)）**

**=(4,m)**

**即对于 X→dim\_feature×dim\_samples 比如有 10 个样本，每个样本有 4 个特征，那么输入为 X4×10**

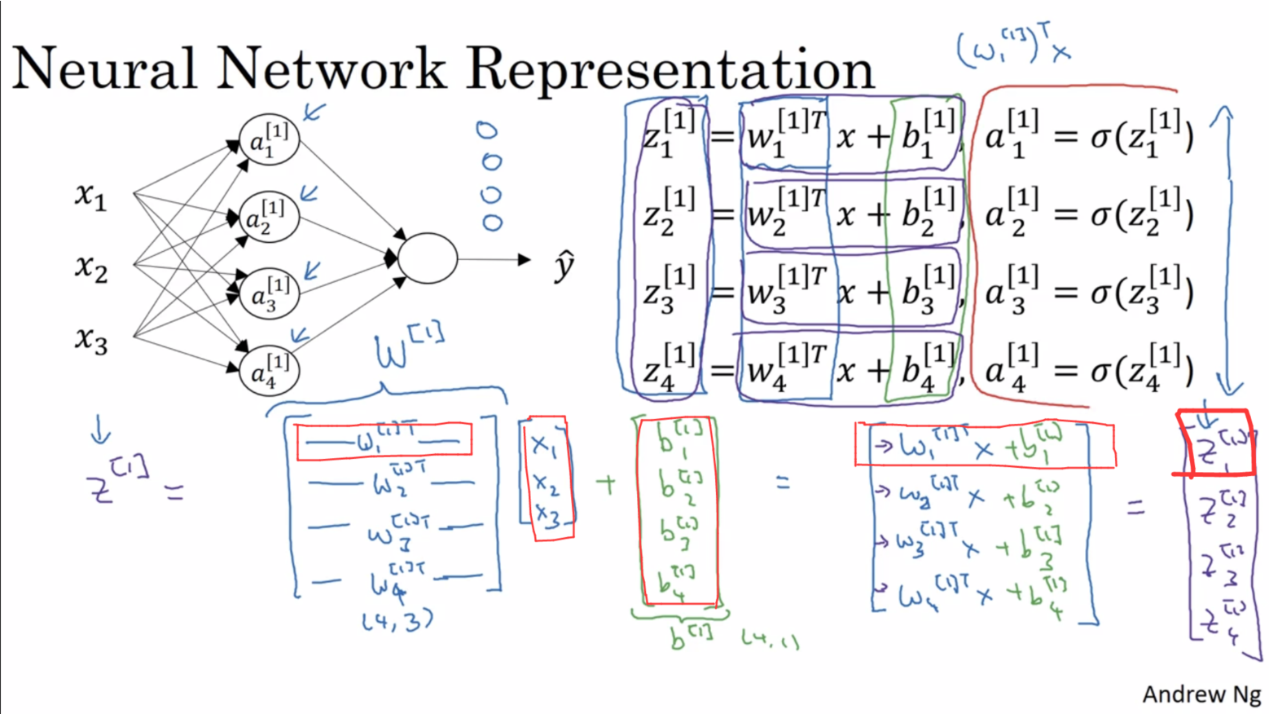
**X=[x1**

**X2**

**X3…]**

**那么 W→num\_units(i+1)×num\_units(i) 即 ( 后一层神经元数 ，前一层神经元数 )对应上图的就是 W3×4**

**W3×4⋅X4×10→matrix3×10**

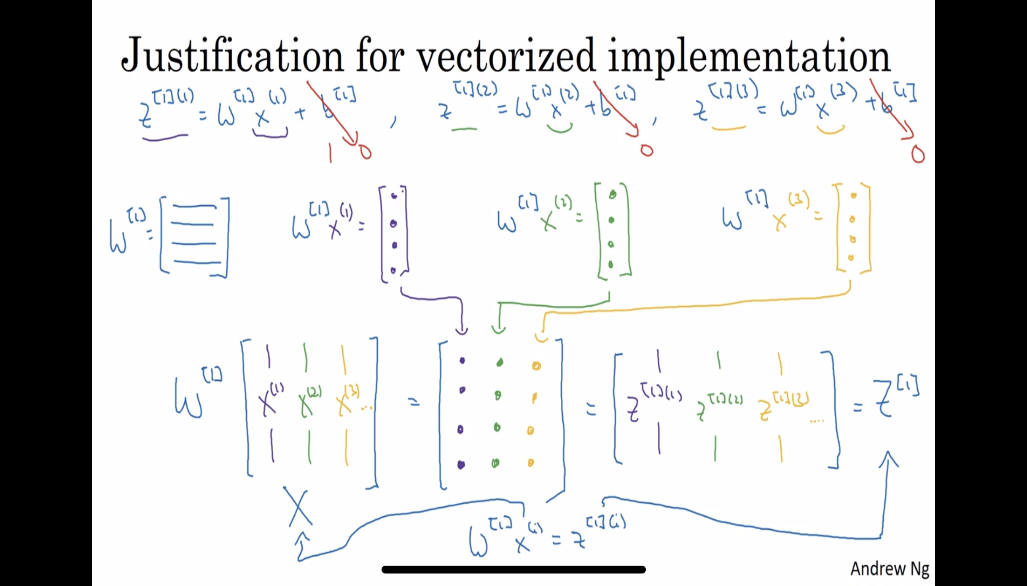


Z【1】=W[1]X+b的过程

=(4，3)\*（1，3）+（4，1）

（4，3）中的第一行 （1，3）中的第一列 +（4，1）中的第一列，得到Z【1】

由此推：

A【1】=σ（WX+b）=σ（Z）

Z[1]=（n【1】,n[0]）T \*(3,1) +(n[1],1)

=（4，3）T\*（3，1）+（3，1）