

**前向传播过程**：

维度：X—2x1；—2x1；W0—2x3；W1—3x2；h—3x1

，维度：3x1。补充：Z0是X通过矩阵相乘对来的，这里可以有，这个结果矩阵成为Jacobina matrix，它是Z0的每一项对X的每一项进行求导，所以Z0有3个占3行，X有2个占两列，求导结果就是

，维度：3x1

，维度：2x1

，维度：2x1

，维度：1x1

**反向传播过程（链式法则）**：

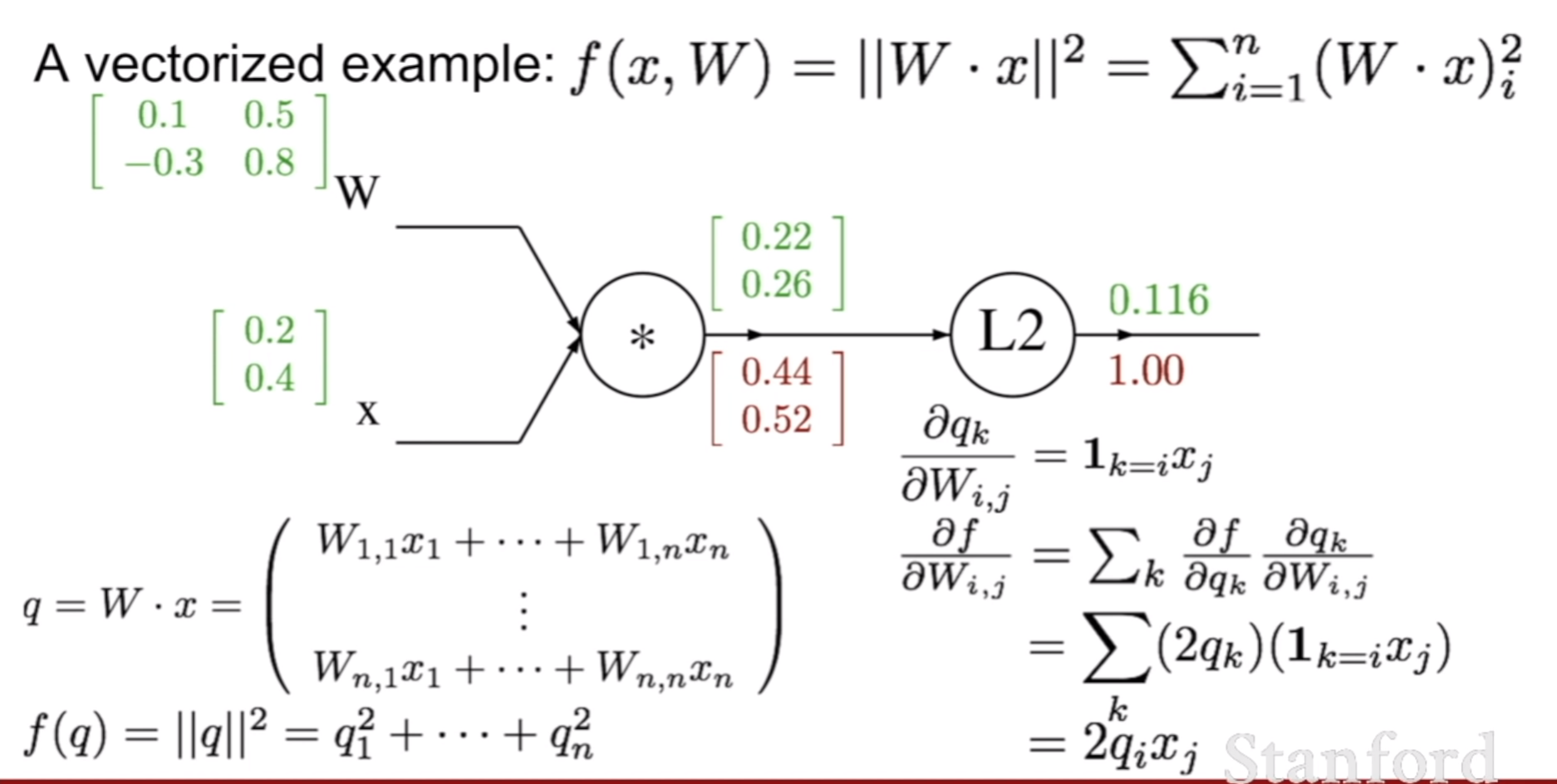
，维度：究竟是2x1还是1x2？按照标量对矢量的求导，理论上返回的结果应该是矢量的转置，如2x1转化成1x2。但是CS231n又说向量梯度的维度应该与原始向量的维度一致，所以这里应该是2x1？**2x1更准确，因为这里维度一致的话，梯度每一项的值就是用来表示对应原始向量值对输出的变化影响程度有多大。**

，维度：3x2

其中，

把带入上式求导，得到

，维度3x2x2。的维度为2x1，为了跟相乘把第三维去掉，这里只能理解为向量内积。所以，如下有

参考这个求

**Always check: the gradient w.r.t a variable should have the same shape as the variable。这里指的都是以L输出为分子，其他变量为分母的偏导。**

，维度2x3

，维度2x1;，维度2x3； ，维度2x1；

，这是错误的。虽然3x1的Z0通过ReLU变成3x1的h，但是这个求导结果不是单值1或者0。从表达式可以看出，是向量对向量的求导，所以这里的结果是一个Jacobian matrix，维度为3x3，即h的每一项都对Z0的每一项进行求导。这样的话，上面求梯度的链式公式就合理。而这个雅各比矩阵实际上是对角阵，其他元素为0，对角线元素不一定全为1，也可能为0。