

每个maxout 单元现在由 k 个权重向量来参数化，而不仅仅是一个，所以maxout 单元通常比整流线性单元需要更多的正则化。如果训练集很大并且每个单元分得的块数保持很低的话，它们可以在没有正则化的情况下正常工作 (Cai *et al.*, 2013)。

maxout 单元还有一些其他的优点。在某些情况下，需要更少的参数可以获得一些统计和计算上的优点。具体来说，如果由 n 个不同的线性过滤器描述的特征可以在不损失信息的情况下，用每一组 k 个特征的最大值来概括的话，那么下一层可以获得 k 倍更少的权重数。

② 因为每个单元由多个过滤器驱动，maxout 单元具有一些冗余来帮助它们抵抗一种被称为灾难遗忘 (catastrophic forgetting) 的现象，这个现象是说神经网络忘记了如何执行它们过去训练的任务 (Goodfellow *et al.*, 2014a)。



