

为稀疏初始化 (sparse initialization) 的替代方案, ①每个单元初始化为恰好有 k 个非零权重。这个想法保持该单元输入的总数量独立于输入数目 m , 而不使单一权重元素的大小随 m 缩小。②稀疏初始化有助于实现单元之间在初始化时更具多样性。③但是, 它也非常偏好于具有很高值的权重。因为梯度下降需要很长时间缩小“不正确”的大值, 这个初始化方案可能会导致单元问题, 例如maxout单元有几个过滤器, 必



