

- ① 如果偏置是作为输出单元，那么初始化偏置以获取正确的输出边缘统计通常是有利的。要做到这一点，我们假设初始权重足够小，该单元的输出仅由偏置决定。这说明设置偏置为应用于训练集上输出边缘统计的激励函数的逆。例如，如果输出是类上的分布，且该分布是高度偏态分布，第 i 类的边缘概率由某个向量 \mathbf{c} 的第 i 个元素给定，那么我们可以通过求解方程 $\text{softmax}(\mathbf{b}) = \mathbf{c}$ 来设置偏置向量 \mathbf{b} 。这不仅适用于分类器，也适用于我们将在第 III 部分遇到的模型，例如自编码器和波尔兹曼机。这些模型拥有输出类似于输入数据 \mathbf{x} 的网络层，非常有助于初始化这些层的偏置以匹配 \mathbf{x} 上的边缘分布。
- ② 有时，我们可能想要选择偏置以避免初始化引起太大饱和。例如，我们可能会将 ReLU 的隐藏单元设为 0.1 而非 0，以避免 ReLU 在初始化时饱和。尽管这种方法不符合不希望偏置具有很强输入的权重初始化准则。例如，不建议使用随机游走初始化 (Sussillo, 2014)。
- ③ 有时，一个单元会控制其他单元能否参与到方程中。在这种情况下，我们有一个单元输出 u ，另一个单元 $h \in [0, 1]$ ，那么我们可以将 h 视作闸门，以决定 $uh \approx 1$ 还是 $uh \approx 0$ 。在这种情形下，我们希望设置偏置 h ，使得在初始化的大多数情况下 $h \approx 1$ 。否则， u 没有机会学习。例如，Jozefowicz *et al.* (2015) 提议设置 LSTM 模型遗忘门的偏置为 1，如第 10.10 节所述。



