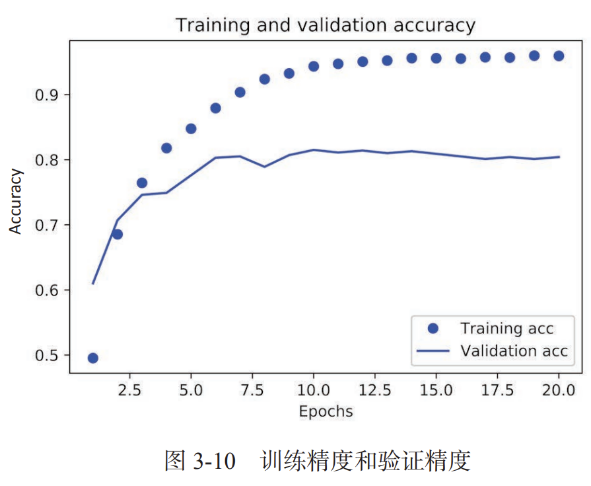
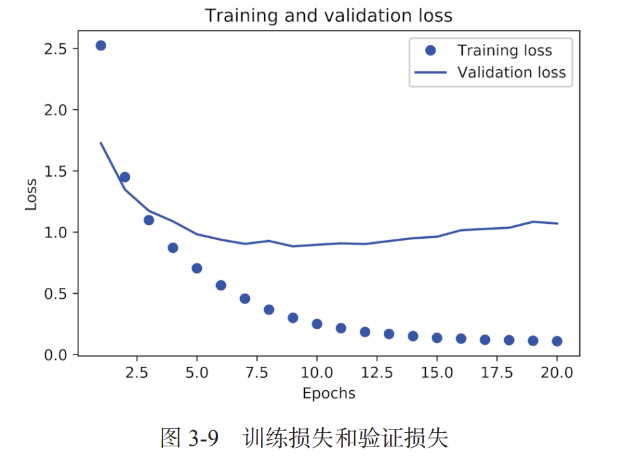
**3.5 单标签多分类问题：路透社新闻主题分类**

****

模型在训练数据上表现得越来越好

但是在验证数据上的性能趋近不变甚至越来越不好

可见在大约第9轮后出现了过拟合

提前在第9轮终止训练可以获得80%左右的精度

相对于完全随机的分类器（50%的精度）表现很好

**信息瓶颈**

目前使用三个Dense层的神经网络进行预测，前两层由64个神经元，最后输出的层有46。

中间的隐藏层的单元不应该比下一层的小太多，不然会造成信息瓶颈。

Layers.dense(64) + Layers.dense(4) + Layers.dense(46)

以上是中间层单元数突然变少的模型

该模型试图将大量信息压缩到维度很小的中间空间

大部分信息能够被塞入中间层 但是并不是全部信息

实验结果表明精度从80%下降到71%

**小结**

* 如果数据分类的类别有N个，网络最有一层应该是大小为N的dense层
* 对于单标签多分类问题，最后一层使用softmax激活，可以输出在每个类别上的概率分布
* 分类问题几乎总是使用分类交叉熵作为损失函数，它使模型输出与真实的概率分布的距离最小化
* 处理多分类问题的标签的两种方法：独热编码 + categorical crossentropy 作为损失函数 或者是 直接将标签存为整数 + sparse categorical crossentropy 损失函数
* 如果数据划分的类别很多，就要避免使用过小的中间层，避免数据穿过中间层是产生信息丢失（信息瓶颈）