注意力加权小例子 成果参考

这个例子解释 softmax 权重如何参与加权，不复现完整 Transformer 训练。

原论文的向量、矩阵投影与多头机制不被简化例子证明；继续阅读原文的缩放与维度定义。

读者问题	解释	边界
权重为什么相加为 1	softmax 除以指数总和	输入得分不同权重也变
5 是模型预测吗	不是，只是标量教学计算	没有运行大模型
论文效果能由此推出吗	不能	需原论文实验

附真实可复算 Python 和可改参数；教材例子不冒充论文实验或 ChatGPT 实际输出。