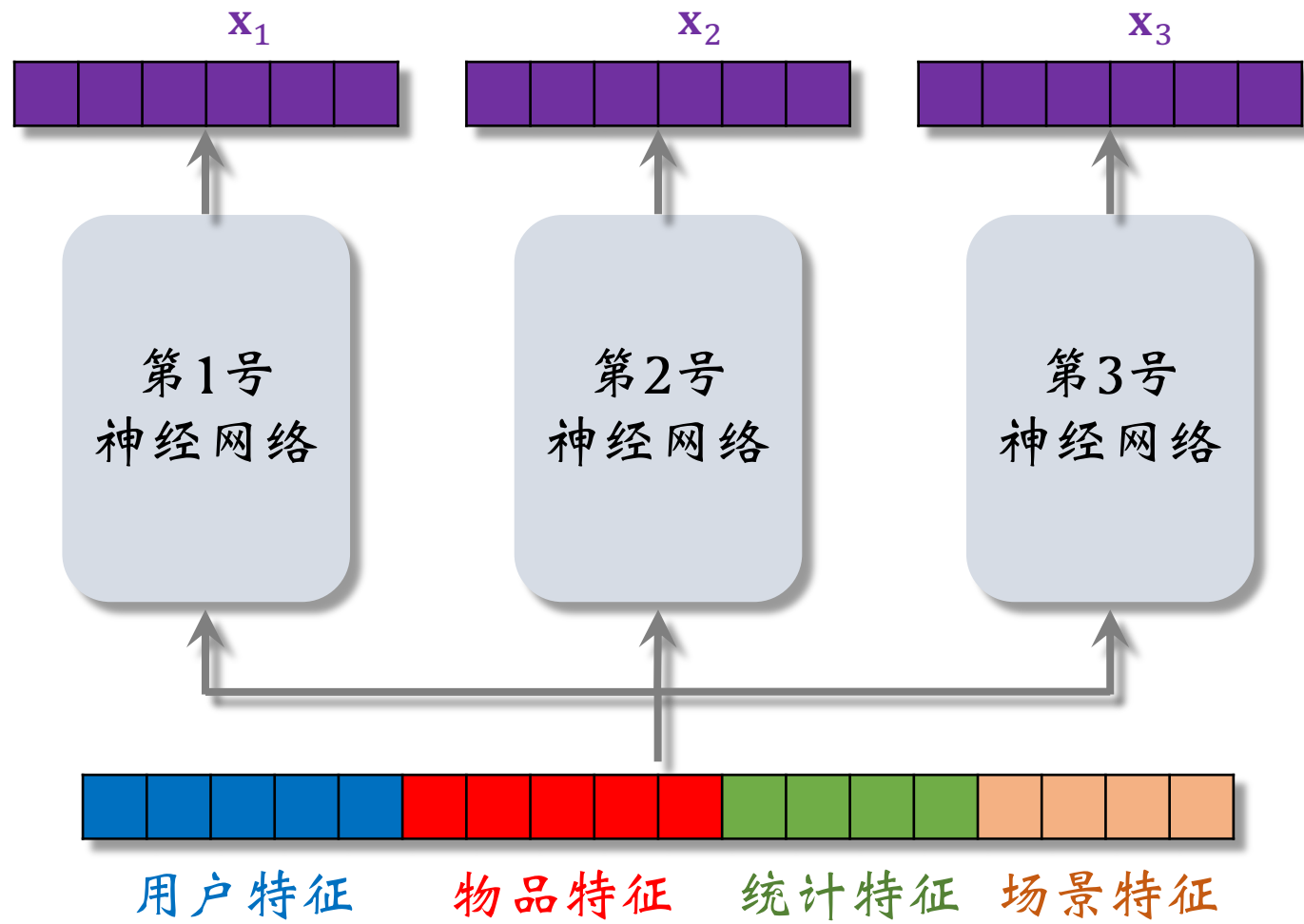


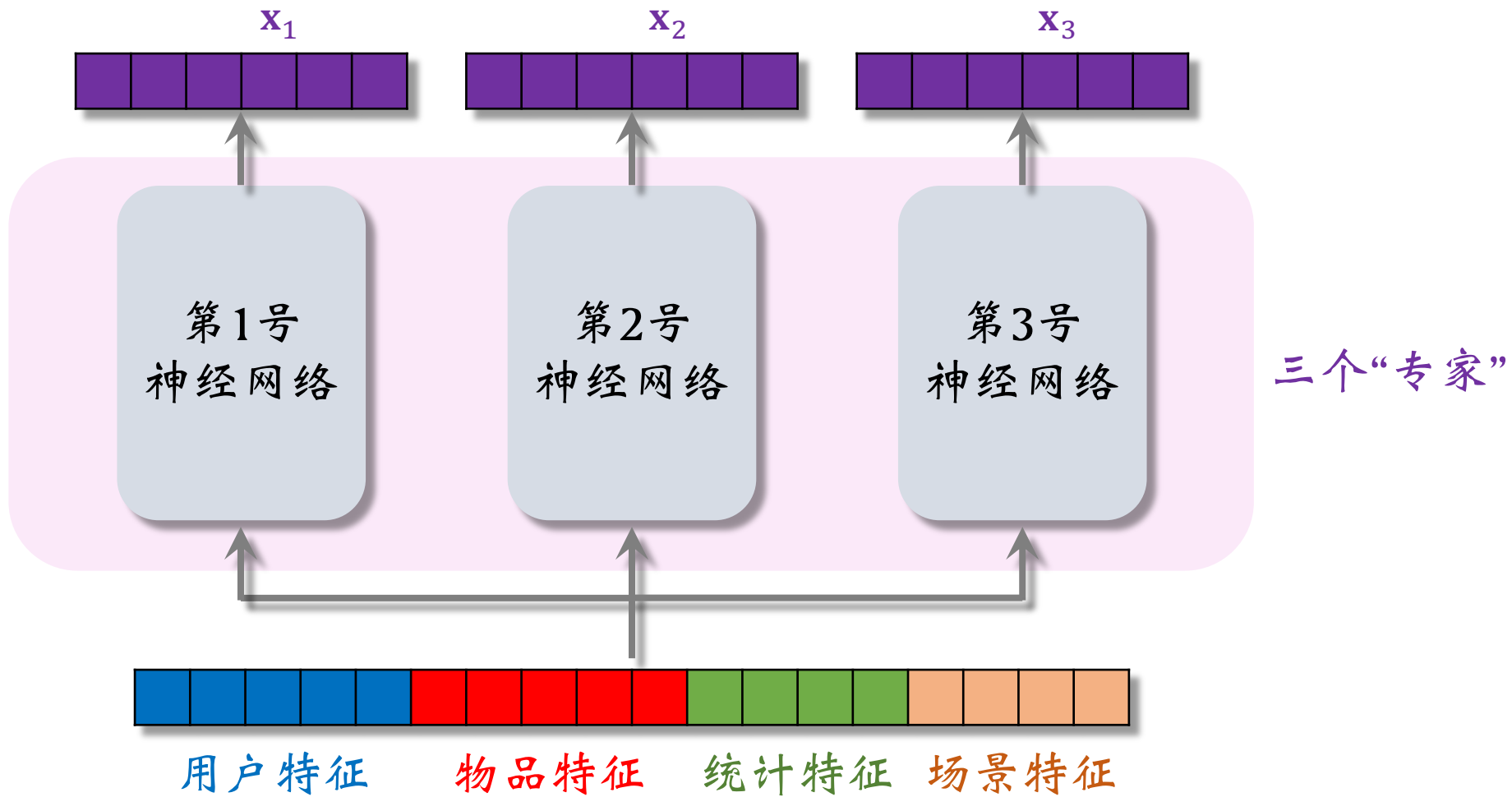
Multi-gate Mixture-of-Experts (MMoE)

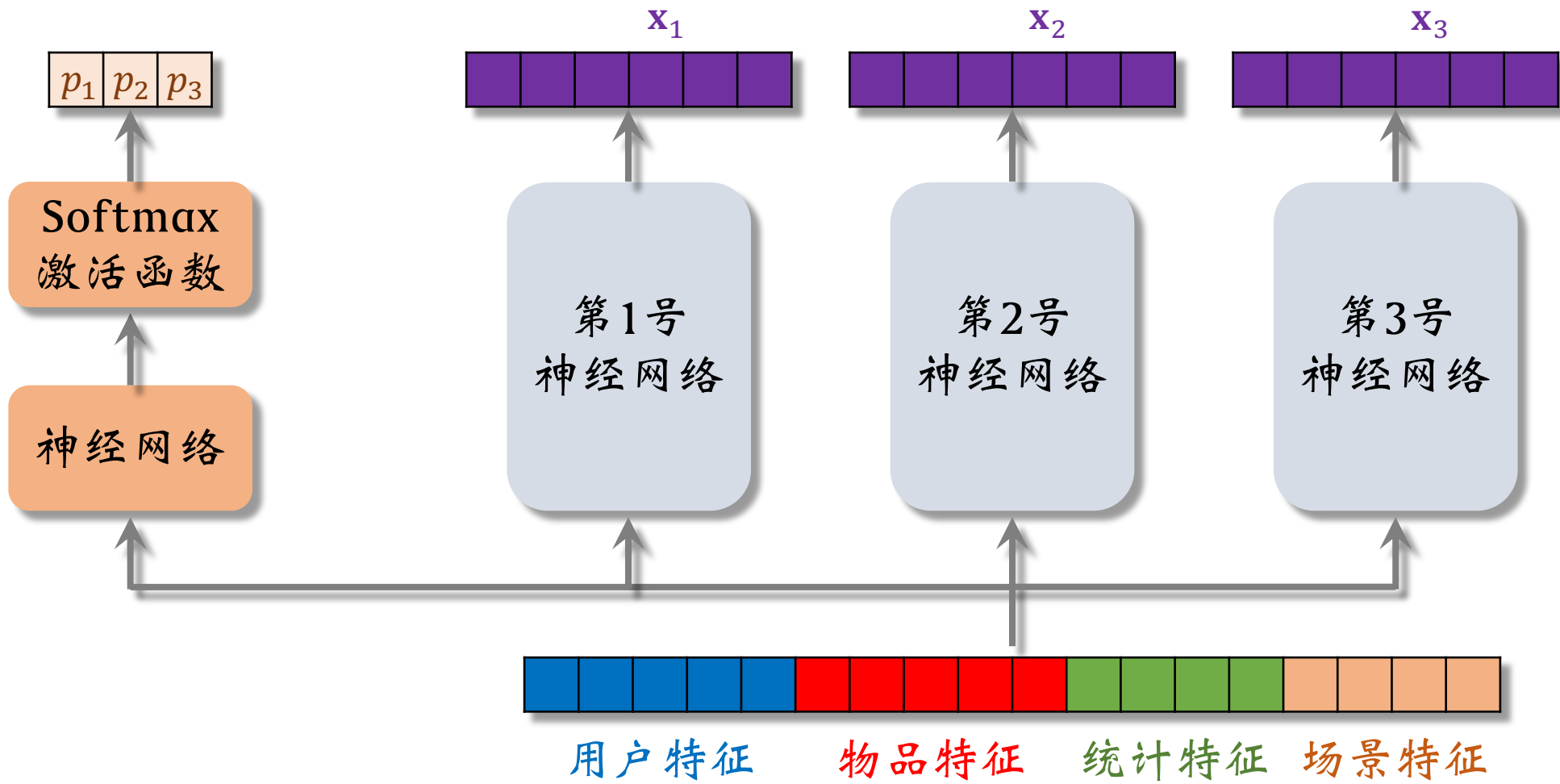
王树森

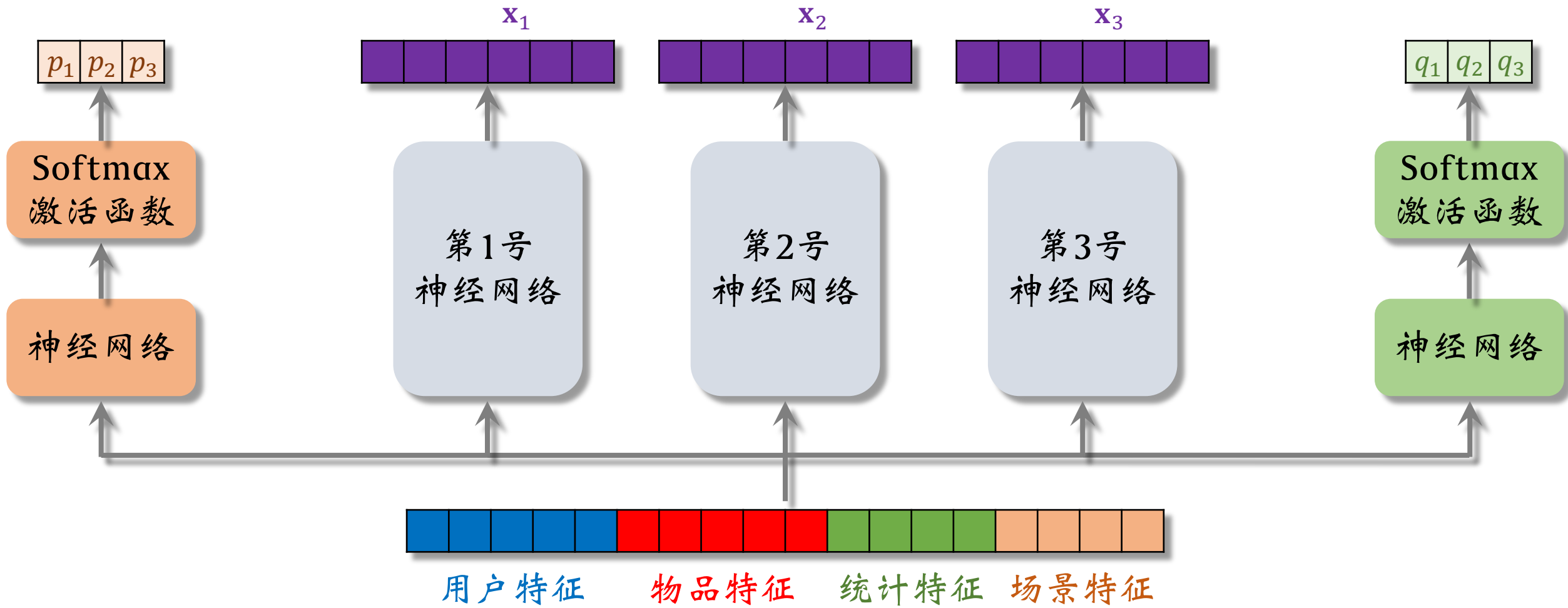
<http://wangshusen.github.io/>

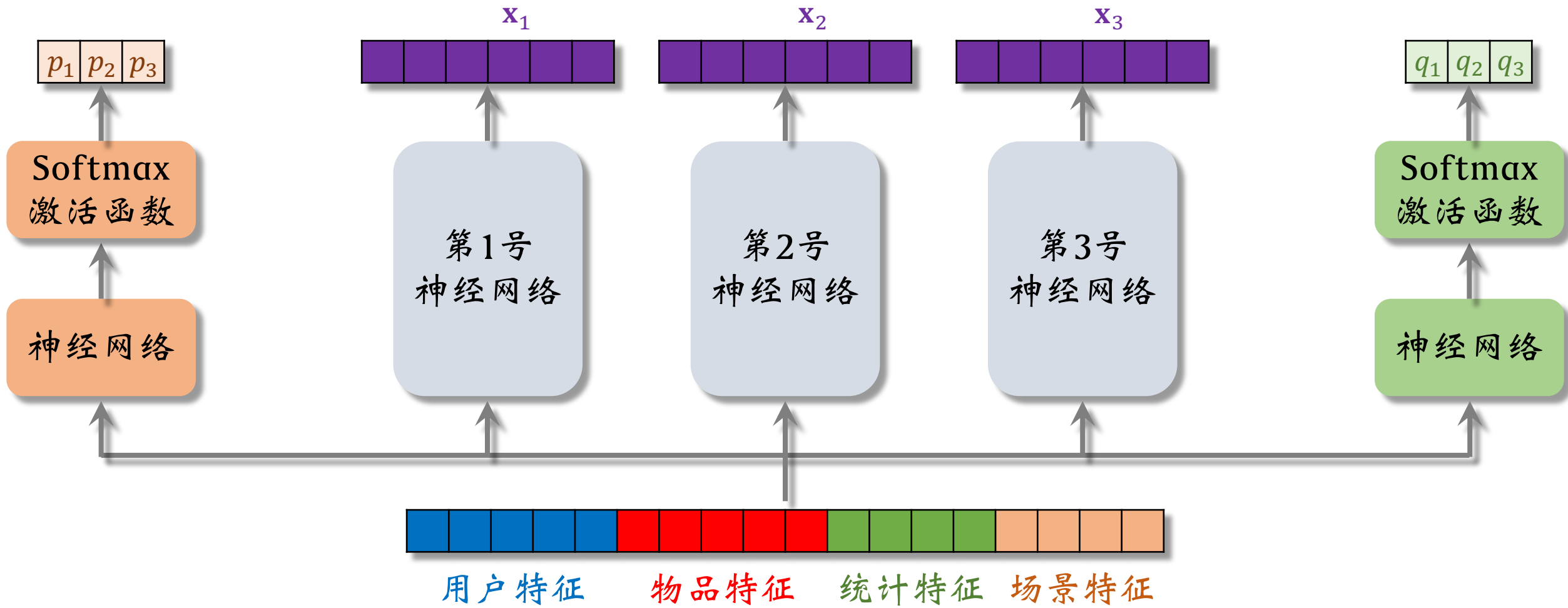












p_1	p_2	p_3
-------	-------	-------

权重

x_1

--	--	--	--	--	--

x_2

--	--	--	--	--	--

x_3

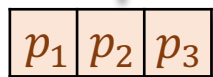
--	--	--	--	--	--

q_1	q_2	q_3
-------	-------	-------

权重



对向量做加权平均



权重

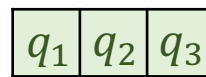
\mathbf{x}_1



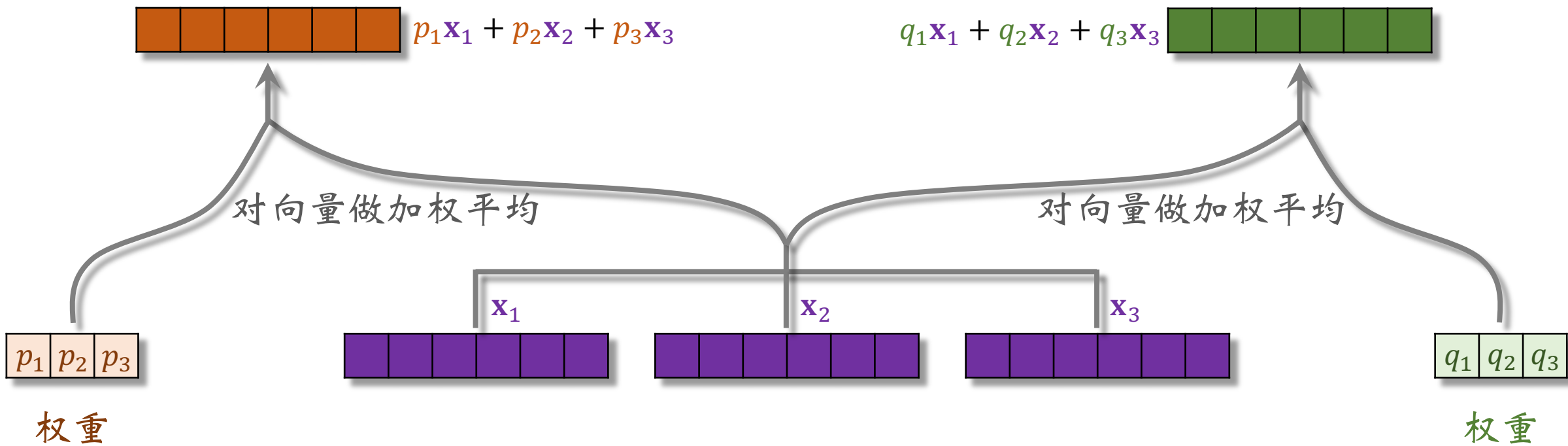
\mathbf{x}_2

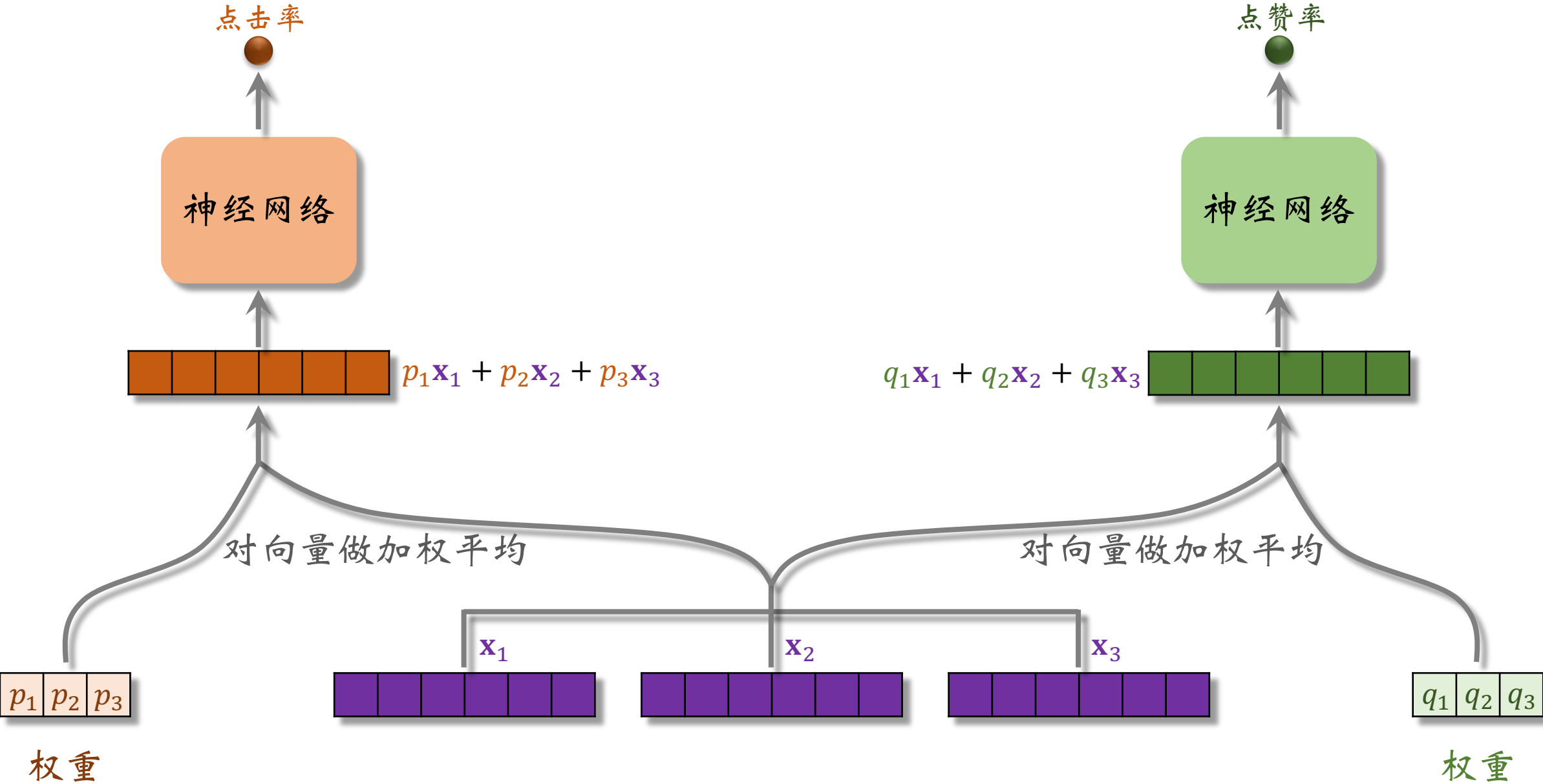


\mathbf{x}_3



权重

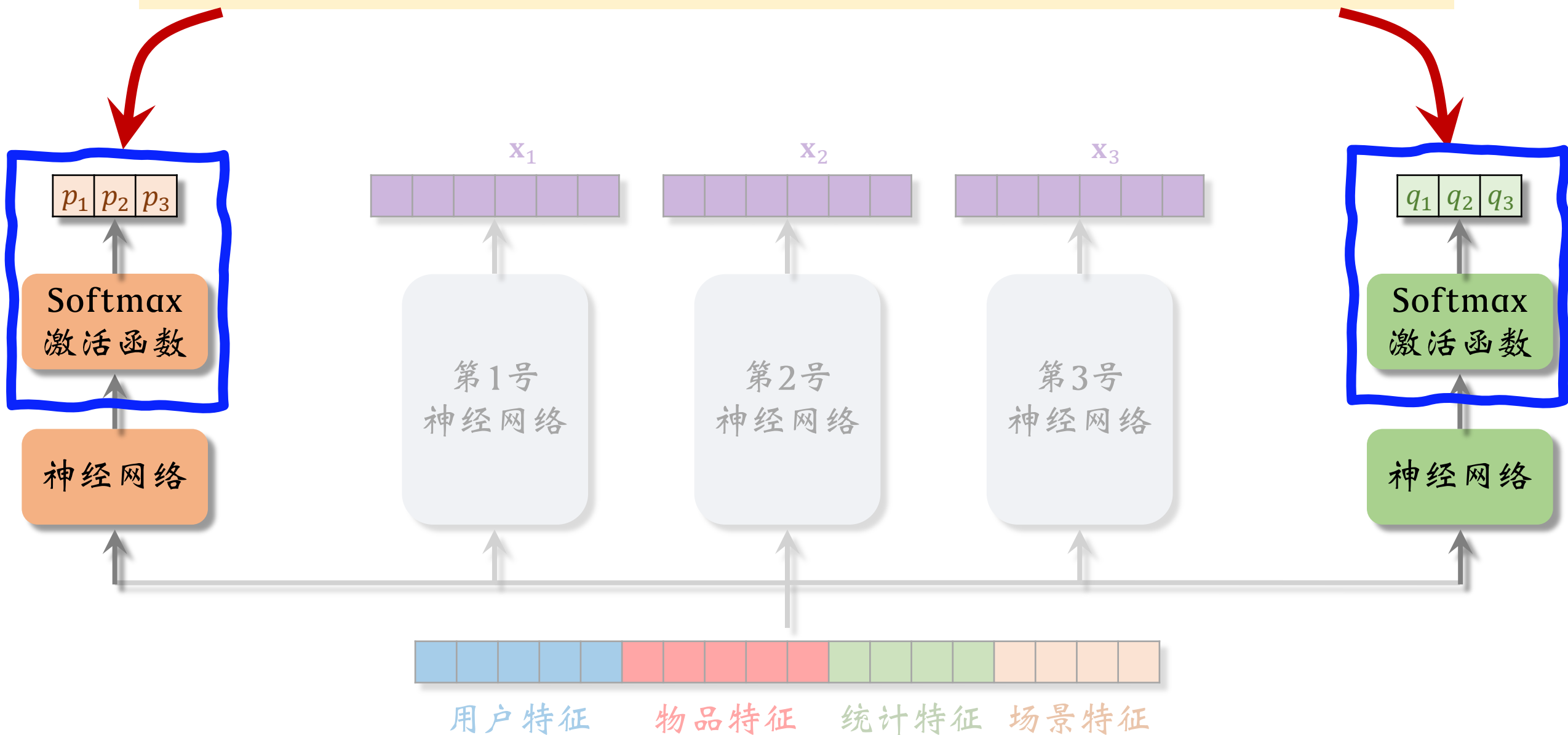




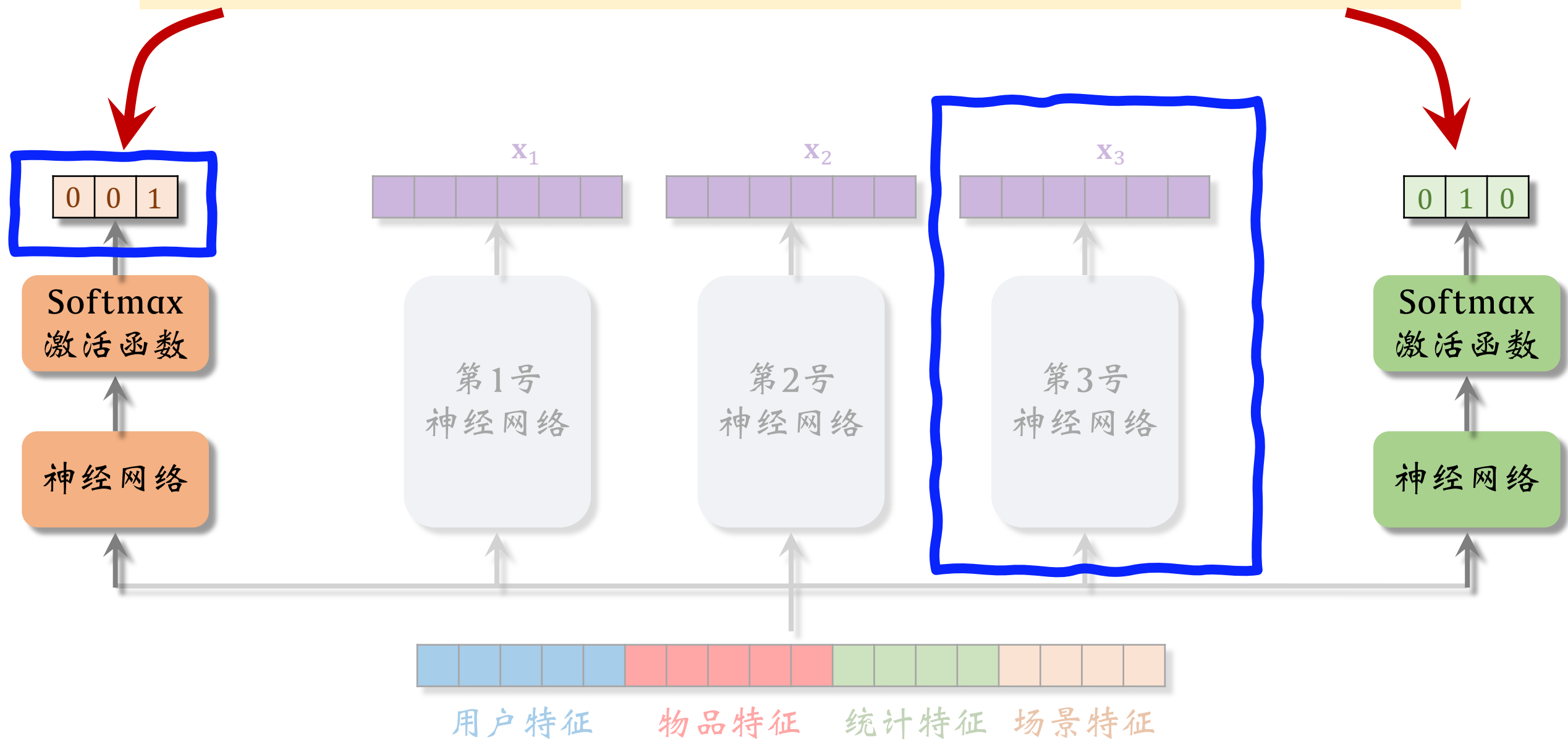
极化现象

(Polarization)

极化(polarize)：Softmax输出值一个接近1，其余接近0。



极化(polarize)：Softmax输出值一个接近1，其余接近0。



解决极化问题

- 如果有 n 个“专家”，那么每个 softmax 的输入和输出都是 n 维向量。
- 在训练时，对 softmax 的输出使用 dropout。
 - Softmax 输出的 n 个数值被 mask 的概率都是 10%。
 - 每个“专家”被随机丢弃的概率都是 10%。

参考文献

- Google 的论文 [1] 提出 MMoE 模型。
- YouTube 的论文 [2] 提出极化问题的解决方案。

参考文献：

1. Jiaqi Ma et al. Modeling Task Relationships in Multi-task Learning with Multi-gate Mixture-of-Experts. In *KDD*, 2018.
2. Zhe Zhao et al. Recommending What Video to Watch Next: A Multitask Ranking System. In *RecSys*, 2019.

Thank You!

<http://wangshusen.github.io/>