Multi-Gate Mixture-of-Experts

Saturday, March 30, 2024   9:52 AM

Multi - Gate Mixture-of-Experts   FE: feature

| user FE | item FE | stats FE | scenario FE |

NOT share parameters across   NNs

NN: neural network

| N N | NN #1 | NN #2 | NN #3 | NN |

Softmax

Expert #1   Expert #2   Expert #3

softmax

$[P_1, P_2, P_3]$ $\vec{X_1}$ $\vec{X_2}$ $\vec{X_3}$ $[\&_1, \&_2, \&_3]$

weighted average   weighted average

$$\vec{x} = P_1 \cdot \vec{X_1} + P_2 \cdot \vec{X_2} + P_3 \cdot \vec{X_3}$$

$$\vec{x} = \&_1 \cdot \vec{X_1} + \&_2 \cdot \vec{X_2} + \&_3 \cdot \vec{X_3}$$

| N N | | N N |

click rate   hike rate
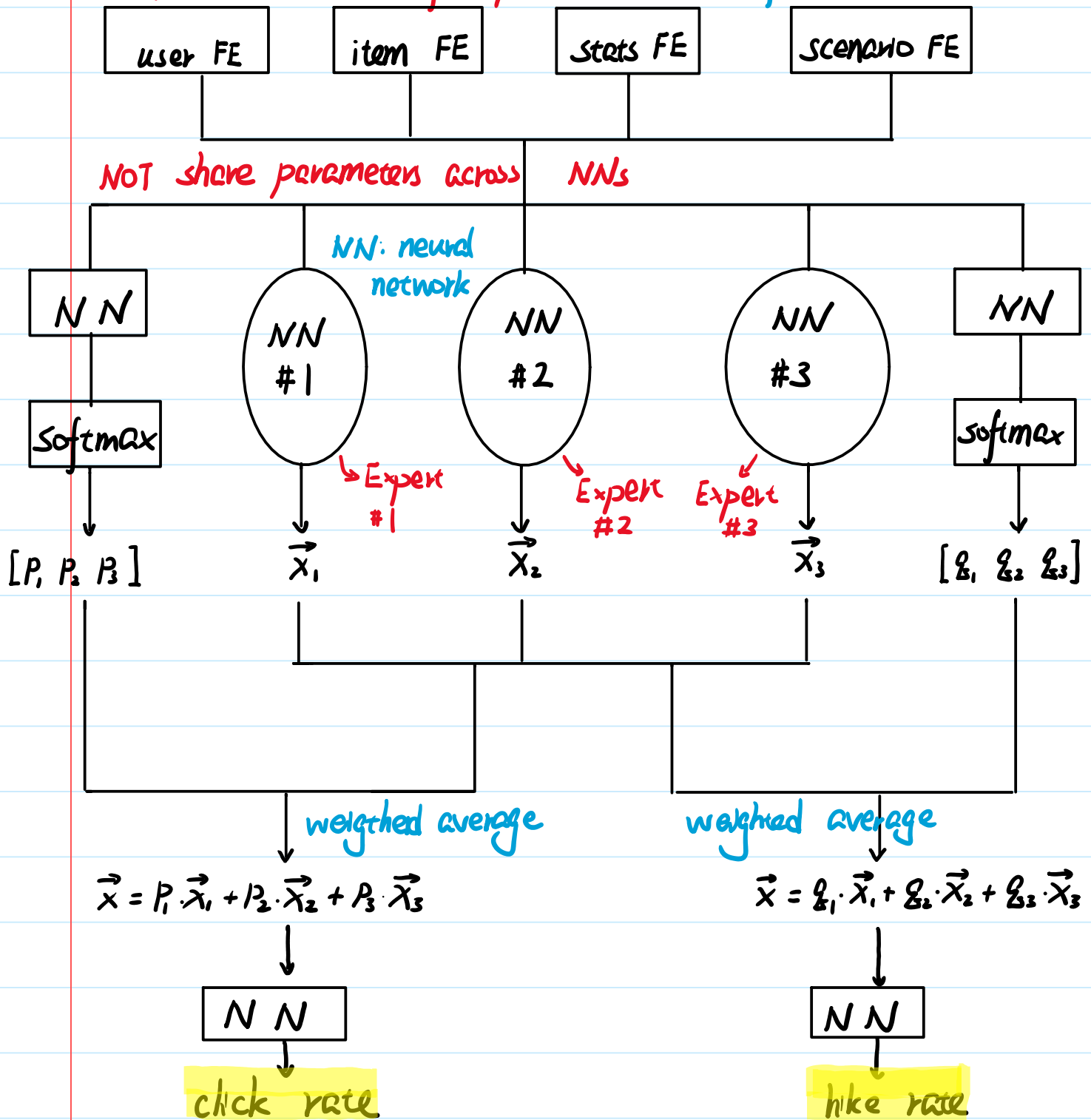
Multi-Gate Mixture-of-Experts

Saturday, March 30, 2024   9:52 AM

Issues :

softmax $\longrightarrow$ [0.01, 0.98, 0.01]

when weighted average, two experts became zero.

Solution :

Drop out :

[P_1 P_2 P_3]

Each element from softmax result has 10% probability to be dropped.

Practice :

① # of experts: 4 ~ 8 (it's 3 in this note)

② May not improve outcome