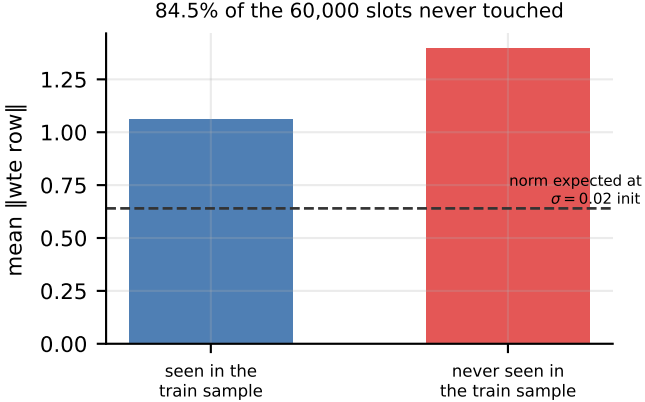
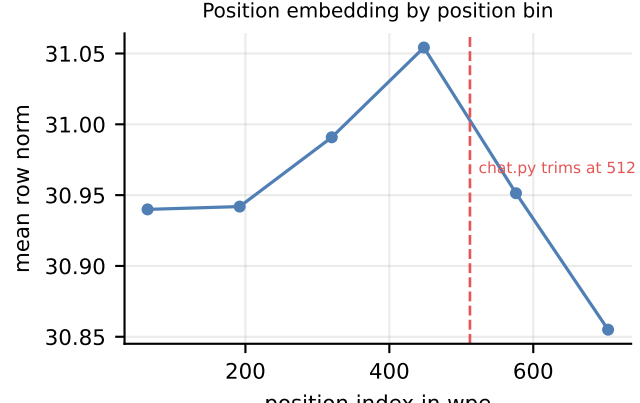

84.5% of the 60,000 slots never touched
mean ‖wte row‖
0.00
0.25
0.50
0.75
1.00
1.25
norm expected at
σ = 0.02 init
seen in the
train sample
never seen in
the train sample


Position embedding by position bin
mean row norm
30.85
30.90
30.95
31.00
31.05
chat.py trims at 512
200
400
600
position index in wpe