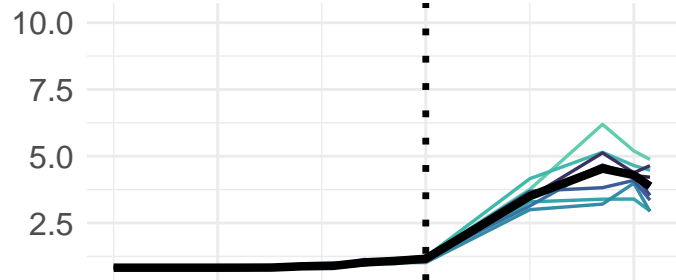
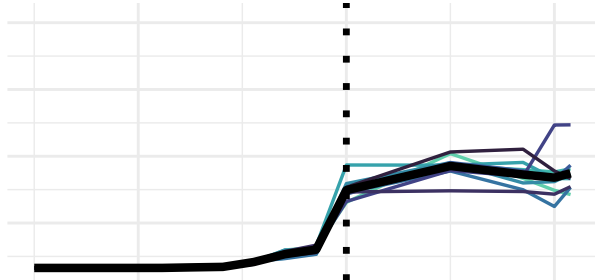


Max Previous/Self Attention Ratio

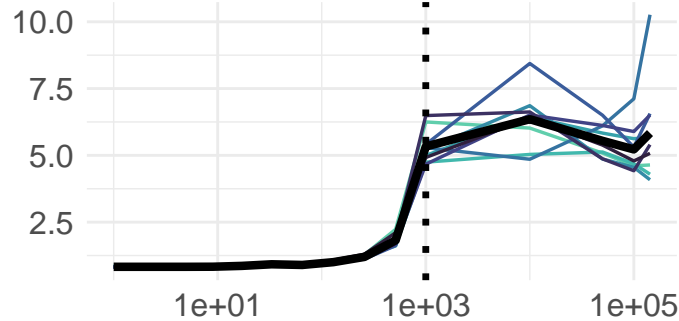
pythia-14m



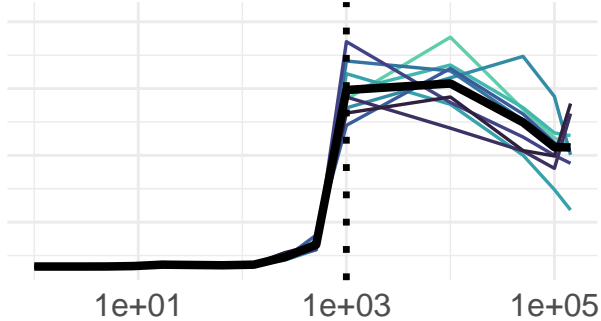
pythia-70m



pythia-160m



pythia-410m



Training Step (Log10)

