

A KL-term Under Broader Exploration

As mentioned in Section 6, our model OURS-B is the sole model to achieve a significant KL-term, yet far from the amount of information one would hope to encode. However, larger KL-terms can be obtained at degraded performance as similarly reported by (Bowman et al., 2016). To illustrate this, we fix a temperature $\beta = 0.065$ and vary β between 0.001 and 10.0. Since α controls the degree of exploration (amount of model generated contexts), the information encoded in the contexts can be varied and the code becomes more important. Similar plots can be obtained for the other methods.

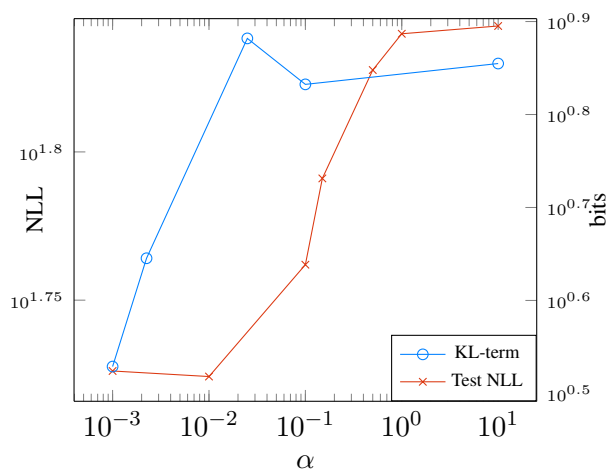


Figure 4: Test NLL and KL-term for various values of α on log-scale.