

In-Context Learning (ICL)

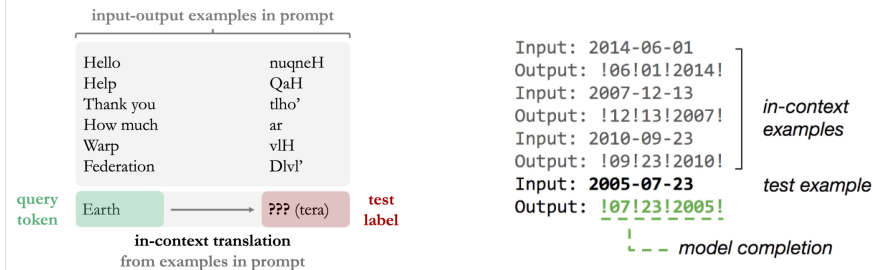


Figure: In-Context Learning / Few-shot Learning

We [Shen et al., 25]² showed that

$$\begin{aligned} & \text{Squared Generalization Error} \\ & \leq \underbrace{C_1 \left(n^{-\frac{2}{2+d}} \log^{1+\frac{3d}{4}} n \right)}_{\text{ability to learn each individual task}} + \underbrace{C_2 \left(\frac{nD^3}{\sqrt{\Gamma}} \sqrt{\log(nD\Gamma)} \right)}_{\text{ability to generalize to new tasks}} \end{aligned}$$

²Zhaiming Shen, Alex Hsu, Rongjie Lai, Wenjing Liao, Understanding In-Context Learning on Structured Manifolds: Bridging Attention to Kernel Methods. *Proceedings of the 14th International Conference on Learning Representations*, 2026

Experiments

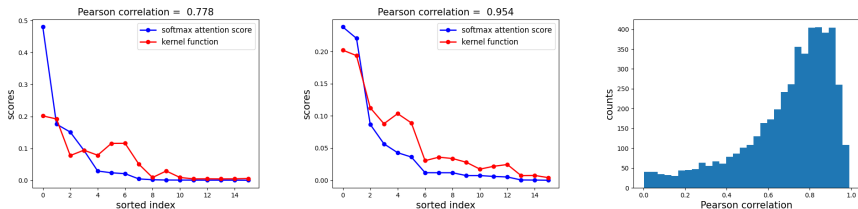


Figure: Left and Middle: attention scores and Gaussian kernel value. Right: histogram of Pearson correlation

Table: Average Pearson correlation and p -value for different prompt length

prompt length n	Pearson correlation	p -value
4	0.86 ± 0.21	0.14 ± 0.21
8	0.75 ± 0.22	0.09 ± 0.19
16	0.69 ± 0.22	0.06 ± 0.17
32	0.67 ± 0.19	0.03 ± 0.12