Beyond Log Likelihood: Probability-based Objectives for Supervised Fine-tuning across the Model Capability Continuum

Gaotang Li*, Ruizhong Qiu*, Xiusi Chen*, Heng Ji[1], Hanghang Tong[1]
[1]University of Illinois Urbana-Champaign

ICML International Conference On Machine Learning 2026

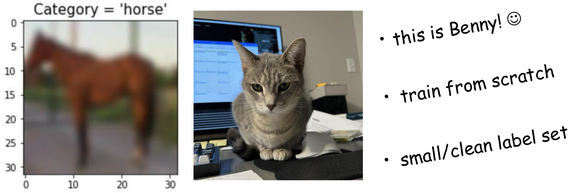
BLENDER IDEA

Paper | Code | Project Page

Main Research Question & Background

What is the right **loss function** for supervised fine-tuning?

Conventional Supervised Learning



- this is Benny! ☺
- train from scratch
- small/clean label set

Language Model SFT

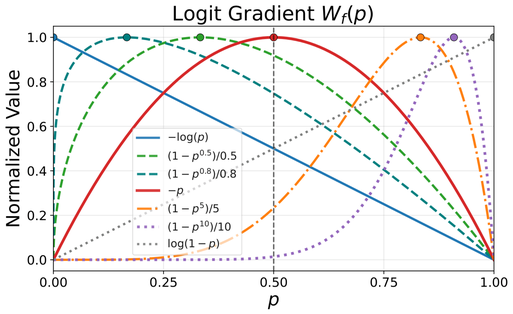
```
<think>
Alright, so I'm trying to solve this problem where I need to find a positive integer \(
    n \) such that \( S_n \), the minimum value of the sum \( \sum_{k=1}^n \sqrt{(2k
    -1)^2 + a_k^2} \), is an integer. Here, \( a_1, a_2, \ldots, a_n \) are positive
    real numbers that add up to 17.

Hmm, okay, let me break this down. I know that when you have a sum of square roots like
    this, it often relates to geometry, maybe minimizing distances. It reminds me of
    the problem where you have to break up a path into segments with certain slopes to
    minimize total distance. Maybe this is similar to finding the shortest path in a
    staircase.
```

- long CoT traces
- pretrained model
- large vocabulary size, noisy labels

Default SFT Loss: Negative Log Likelihood (NLL) $-\log(p)$ p: predicted probability of imitation label (token)

Q: What is so special about $-\log(p)$ and what about other probability-based loss function $f(p)$?



Prior-leaning vs. Prior-averse Objectives

- How much learning signal each token contributes relative to model's prior belief
- learning signal: gradient norm
- Prior belief: predicted probability of target token

Lemma 1 (Gradient Shape). *Let $f : [0,1] \to \mathbb{R}$ be differentiable and nonincreasing. Then the gradient of Eq. 2 with respect to the logits at step t is*

$$\frac{\partial(\mathcal{L}_f)}{\partial z_{t,i}} = s_f(p_{t,y})\,(\delta_{i,y} - p_{t,i}), \qquad \text{where } s_f(p) \triangleq -f'(p)\,p \geq 0,\ \delta_{iy} = 1\{i = y\}.$$

In particular, for the correct class $i = y$,

$$\frac{\partial(\mathcal{L}_f)}{\partial z_{t,y}} = s_f(p_{t,y})\,(1 - p_{t,y}) = W_f(p_{t,y}), \qquad W_f(p) \triangleq -f'(p)\,p\,(1-p).$$

$$\mathcal{L}_{f(p)}(\theta) = \mathbb{E}_{(x,\tilde{y})\sim T}\left[f(p_\theta(\tilde{y}|x))\right]$$

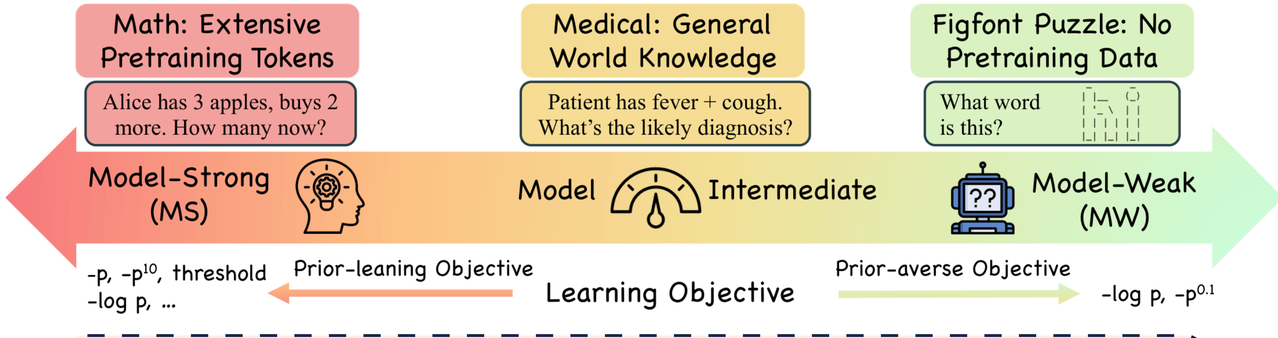
$$= \mathbb{E}_{(x,\tilde{y})\sim T}\left[\sum_{t=1}^{N} f(p_\theta(y_t|y_{<t}, x))\right]$$

p_θ:model output (probability vector); token t logit vector $z_t = (z_{t,1}, \ldots, z_{t,V})$

$\tilde{y}$: SFT token; x: prompt; t: token position; i: the i-th class in logit vector ; token t logit vector correct class $z_{t,y}$

Examples of some choice of f(p):
(1) $-\log(p)$: prior-averse. (2) $-p$: prior-leaning (3) $(1-p^{\alpha})/\alpha$: a general-form

Main Findings



We uncover a **model-capability continuum**, which characterizes different SFT objectives

- ❖Model-Strong: Extensive Pretraining Data, Strong Model Prior, Prior-leaning objectives wins
- ❖Model-Intermediate: Partial World Knowledge Intermediate Model Prior, No Obvious winner
- ❖Model-Weak: No/Limited Pretraining Data, Weak Model Prior, Prior-averse Objective wins

Experiments and theory

- ❖Validate the continuum view across 8 model backbones, 27 benchmarks, and 7 domains
- ❖Extensive ablation studies with different objectives and explicit quantile threshold
- ❖Theoretical characterization of the model-strong and model-weak end

Checkout the paper for more details!