Deterministic Bounds and Random Estimates of Metric Tensors on Neuromanifolds

A neural classifier $p(y \mid x, \theta)$ with weights $\theta \in \Theta$.

$\Theta = \{\theta\}$: Neuromanifold

Geometry of Θ is induced by the Fisher information matrix (FIM):

$$F(\theta) = \sum_{x \in D_x} \mathbb{E}_{p(y|x,\theta)} \left[\frac{\partial \log p(y \mid x, \theta)}{\partial \theta} \frac{\partial \log p(y \mid x, \theta)}{\partial \theta^\top} \right]$$

Fundamental to optimization, pruning, transfer learning, and learning theory.

Empirical Fisher

$$\overline{F}(\theta) = \sum_{(x,y)\in D} \left[\frac{\partial \log p(y \mid x, \theta)}{\partial \theta} \frac{\partial \log p(y \mid x, \theta)}{\partial \theta^{\mathsf{T}}} \right].$$

Widely used; biased; not guaranteed to recover $F(\theta)$.

Monte Carlo Estimate

$$\hat{F}(\theta) = \frac{1}{m} \sum_{k=1}^{m} \frac{\partial \log p(\hat{y}_k \mid \hat{x}_k, \theta)}{\partial \theta} \frac{\partial \log p(\hat{y}_k \mid \hat{x}_k, \theta)}{\partial \theta^{\mathsf{T}}}$$

$$\hat{x}_k \sim \mathrm{Uniform}(D_x), \;\; \hat{y}_k \sim p(y \mid \hat{x}_k, \theta).$$

Unbiased; variance can be unbounded; expensive.

Hutchinson FIM

$$h(\theta) = 2 \sum_x \sum_{y=1}^{C} \sqrt{p(y \mid x, \theta)}\ \ \xi_{xy}$$

ξ_{xy} are independent Rademacher (±1) variables.

A single backward pass gives $\frac{\partial h}{\partial \theta}$.

Its outer product estimates the FIM:

$$\mathbb{F}(\theta) = \frac{\partial h}{\partial \theta} \frac{\partial h}{\partial \theta^\top}$$

Unbiased — Hutchinson cancels cross terms in expectation.

Bounded CV — upper-tail control for the FIM estimate.

Cheap — one backward pass.

$\mathbb{F}(\theta)$ remains scalable for modern deep classifiers and invites broader applications.