Title:

Entropy Structure of Scaling-Normalized Vector Fields and Its Stochastic Interpretation via Itô Calculus

Fukiya,Kazuo

Abstract:

This paper investigates the entropy structure of scaling-normalized vector representations in high-dimensional language spaces. We show that high-entropy vectors are typically distributed near the origin, while low-entropy vectors tend to extend along basis directions and lie far from the origin. This geometric property is analyzed in connection with stochastic differential equations (SDEs) and Itô calculus. In particular, we apply Itô's lemma to derive a stochastic formulation that characterizes semantic drift and diffusion along the vector field, providing a probabilistic interpretation of entropy gradients in language models.

1. Introduction

Language models represent words or phrases as high-dimensional vectors. To analyze these vectors in a geometrically meaningful way, we apply *scaling normalization*, which preserves directional features while projecting the vector field onto a bounded norm space.

Interestingly, the entropy of such normalized representations reveals a clear spatial pattern: high entropy vectors concentrate near the origin, whereas low entropy vectors align along the principal basis vectors at a distance.

This study explores this phenomenon both geometrically and probabilistically, leveraging the machinery of stochastic differential equations (SDEs), Itô integrals, and entropy geometry.

2. Scaling Normalization and Entropy

Let $x \in \mathbb{R}^n$ denote a word vector. We define the scaling-normalized vector as:

$$\hat{x} = \frac{x}{\| x \|}$$

where $\|\cdot\|$ is the Euclidean norm. The entropy $H(\hat{x})$ is considered with respect to the probability density of $\hat{x}$ over the unit ball. Empirically, we observe:

- $H(\hat{x})$ is *high* when $\| x \|$ is small, meaning vectors lie near the origin.
- $H(\hat{x})$ is *low* when $\| x \|$ is large and aligned with a canonical basis.

This suggests an *entropy field* over the vector space, where semantic specificity (low entropy) is linked to directionality and magnitude.

3. Stochastic Representation of Vector Dynamics

To model the evolution of these vectors, we consider a stochastic process X_t governed by the SDE:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$$

where μ is a drift term, σ is a volatility matrix, and W_t is an n-dimensional Wiener process.

We apply **Itô's lemma** to a scalar function $f(X_t)$, e.g., entropy or norm function:

$$df(X_t) = \left(\nabla f \cdot \mu + \frac{1}{2}Tr[\sigma^T \nabla^2 f \sigma]\right)dt + \nabla f \cdot \sigma dW_t$$

This allows us to compute how the entropy (or its proxy) changes as vectors diffuse in the space. In particular, when entropy is viewed as a scalar field $H(X)$, we model its local behavior through:

$$dH(X_t) = \mathcal{L}H(X_t)dt + \nabla H \cdot \sigma dW_t$$

where $\mathcal{L}$ is the generator of the diffusion process.

---

4. Interpretation in Language Models

In language representation spaces, vector movement due to contextual updates (e.g., during prediction or pretraining) can be seen as a stochastic trajectory. High entropy regions near the origin reflect *semantic ambiguity* or *polysemy*, whereas low entropy vectors away from the origin represent *semantic sharpness* or *specificity*.

Thus, the entropy gradient $\nabla H(x)$ can guide attention or sampling in generation models.

---

5. Conclusion and Future Work

We have presented a framework that links the geometric distribution of normalized language vectors with entropy and stochastic dynamics. The application of Itô calculus provides a deeper understanding of how vectors evolve and encode meaning probabilistically.

Future work will include the empirical estimation of drift and diffusion terms from transformer-based models and the construction of entropy-aware sampling algorithms.
