The Kac-Rice Formula for Landscape Complexity

Slides by Ian Joffe

Based on *Landscape Complexity for the Empirical Risk of Generalized Linear Models* by Antoine Maillard, Gerard Ben Arous, and Giulio Biroli, 2019

Presented February 2024 for Stat 37797

Setup

We consider random GLMs

$$y_\mu = \phi(\xi_\mu \cdot x^*)$$

- ξ_μ is a gaussian random vector
- x^* the the true vector we are trying to recover, $||x^*||_2^2 = 1$
- ϕ is a nonlinearity
- There are m samples, indexed by μ
- x^* is n-dimensional
- Ex: Phase Retrieval

We estimate x^* with x. The loss is

$$L_2(x) = \frac{1}{2m}\sum_{\mu=1}^{m}[\phi(\xi_\mu \cdot x) - y_\mu]^2$$

How many critical points does L_2 have?

Understanding Kac-Rice

Kac-Rice counts the number of critical points, N, of a function by counting the zeros of its derivative over some interval. Let γ be the curve characterized by L, and let $\delta(\cdot)$ be the dirac delta.

$$\begin{aligned}
N &= \int_\gamma 1 \cdot \delta(\omega) d\omega \\
&= \int_a^b \delta(L'(x)) |L''(x)| dx \\
\mathbb{E}(N) &= \int_{x=a}^b \int_{y=-\infty}^\infty \mathbb{E}\left[\delta(L'(x)) |L''(x)| \middle| L'(x) = y \right] p_{L'(x)}(y) dy dx \\
&\qquad \text{(conditioning on } L'(x)\text{)} \\
&= \int_{x=a}^b \int_{y=-\infty}^\infty \delta(L'(y)) \mathbb{E}\left[|L''(x)| \middle| L'(x) = y \right] p_{L'(x)}(y) dy dx \\
&= \int_a^b \mathbb{E}\left[|L''(x)| \middle| L'(x) = 0 \right] p_{L'(x)}(0) dx
\end{aligned}$$

Understanding Kac-Rice

$$\mathbb{E}(N) = \int_a^b \mathbb{E}\left[|L''(x)| \middle| L'(x) = 0 \right] p_{L'(x)}(0) dx$$

In multiple dimensions, infinitesimal distances are scaled by $|\det \nabla(\nabla L)| = |\det \nabla^2 L|$

The Multidimensional Kac-Rice Formula

$$\mathbb{E}(N) = \int_A \mathbb{E}\left[|\det(\nabla^2 L(x))| \middle| \nabla L(x) = \mathbf{0} \right] p_{\nabla L(x)}(\mathbf{0}) dx$$

This understanding from Tim Kunisky's course CPSC 664

Applying Kac-Rice

The landscape complexity expresses the number of critical points

- Annealed complexity: $\log \mathbb{E}(N)$
- Quenched complexity: $\mathbb{E}(\log N)$
- Thermodynamic limit: $n, m \to \infty; m/n \to \alpha > 1$

Let y_μ and z_μ be independent Gaussian random vectors. Using $L = \frac{1}{m}\sum_{\mu=1}^{m} \phi(\xi_\mu \cdot x)$,

$$L(x) \overset{d}{=} \frac{1}{m}\sum_{\mu=1}^{m} \phi(y_\mu)$$

$$\nabla L(x) \overset{d}{=} \frac{1}{m}\sum_{\mu=1}^{m} \phi'(y_\mu) z_\mu$$

$$\nabla^2 L(x) \overset{d}{=} \left[\frac{1}{m}\sum_{\mu=1}^{m} \phi''(y_\mu) z_\mu z_\mu^T\right] - \left[\left(\frac{1}{m}\sum_{\mu=1}^{m} y_\mu \phi'(y_\mu)\right) I_{n-1}\right]$$

∇L and $\nabla^2 L$ do not depend on x!

Applying Kac-Rice

Since $\nabla L, \nabla^2 L \perp x$, the integral is a "rectangle" and we can evaulate at any point, say $x = [1, 0, 0, ..., 0]^T = e_n$. Using Gaussian calculations,

$$
\begin{aligned}
\mathbb{E}(N) &= \int_A \mathbb{E}\left[|\det(\nabla^2 L(x))| \,\middle|\, \nabla L(x) = \mathbf{0} \right] p_{\nabla L(x)}(\mathbf{0}) dx \\
&= \omega_n \cdot p_{\nabla L(e_n)}(\mathbf{0}) \cdot \mathbb{E}_{y,z}\left[|\det(\nabla^2 L(e_n))| \,\middle|\, \nabla L(e_n) = \mathbf{0} \right] \\
&= \omega_n \cdot \mathbb{E}_y\left[p_{\nabla L(e_n)}(\mathbf{0}) \cdot \mathbb{E}_z\left[|\det(\nabla^2 L(e_n))| \,\middle|\, \nabla L(e_n) = \mathbf{0},\ y \right] \right] \\
&= C e^{n\frac{1+\log\alpha}{2}} \mathbb{E}_y\left[e^{-\frac{n-1}{2}\log(\frac{1}{m}\sum_\mu \phi'(y_\mu)^2)} \mathbb{E}_z\left[|\det H_n^\Lambda(y, z)| \right] \right]
\end{aligned}
$$

C, which mostly comes from ω_n (the volume of A), is trivial at the limit we will take, and H_n^Λ is a messy nxn matrix.

Results at the Limit

For the L we have been working with, large deviation theory (see Varadhan's Lemma) yields annealed complexity

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}(N) = \sup_{\nu \in \mathcal{M}_\phi(A)} \Big[\frac{1+\log\alpha}{2} - \frac{\mathcal{E}_\phi(\nu)}{2} + \kappa_{\alpha,\phi}(\nu, t_\phi(\nu)) - \alpha H(\nu \big| \mu_G) \Big]$$

A similar procedure leads us to

Annealed Complexity of L_2

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}(N) = \frac{1+\log\alpha}{2} + \sup_{q\in Q} \sup_{\nu \in M_\phi(B,q)} \Big[\frac{1}{2}\log(1-q^2) - \frac{1}{2}\mathcal{E}_\phi(q,\nu) + \kappa_{\alpha,\phi}(q,\nu) - \alpha H(\nu \big| \mu_G) \Big]$$

where Q is the set of allowed signal sizes $x^* \cdot x$ ($Q \subseteq [-1,1]$ since $||x||_2^2 = 1$)

Using the Result

M,A,&B develop a fixed point iteration algorithm to compute the annealed complexity from the supremum formulas. For my report, I'm working to use this to get an exact number for N with L_2 loss on phase retrieval.

Augeri, Guionnet, and Husson (2021) used a method called "tilting of the measure" to count local minima (instead of all critical points).