

Роль логарифмів у машинному навчанні

Тарас Серєда
taras.y.sereda@proton.me

Вступ

Машинне навчання – це прикладна дисципліна котра пов’язує лінійну алгебру, теорію ймовірності, математичний аналіз та комп’ютерні науки для побудови математичних моделей, що навчаються з даних.

Теорія ймовірності у нейронних мережах



Figure 1: Дивний кубик створений chatGPT.

Ймовірність події прийнято чисельно описувати значеннями в діапазоні $0 \leq P(x) \leq 1$. Де 0 – не можлива подія, а 1 – гарантована подія. Поняття ймовірності і сюрпризу від настання події є обернено пропорційними. Чим більш ймовірна подія тим менша міра нашого здивування. Оскільки ймовірності деяких подій можуть бути дуже малими, наприклад < 0.00005 . То буде зручним перейти до логарифмічного простору для збереження точності обчислень при множенні малої ймовірних подій. Тоді така подія буде мати логарифмічну ймовірність $\log(0.00005) = -9.90348755$

Приклад. Нехай X є випадкова змінна $X \in \{1, 2, 3, 4, 5, 6\}$ що описує підкидання грального кубика. Для кожного значення ймовірність дорівнює $P(X = x) = \frac{1}{6}$, де x є одним з можливих результатів кубика.

1. $S(x) = -\log P(X = x)$ Сюрприз, описує міру мого здивування.
2. $H(p) = -\sum P(x) \cdot \log P(x)$ Ентропія. Виражає очікувану міру мого здивування. Для випадку з гральним кубиком $S(x) = H(x)$
3. $H(p, q) = -\sum P(x) \cdot \log Q(x)$ Перехресна Ентропія (cross entropy). Виражає моє очікування того на скільки нейронна мережа буде здивована якщо функція розподілу її передбачень є $Q(x)$ а не $P(x)$.

Нейронні мережі тренуються шляхом покрокової оптимізації та обчислення функції втрат, що описує міру невідповідності поточного стану нейронної мережі до тренувальних даних. Саме перехресна ентропія є ключовою функцією втрат, що використовується для навчання популярних моделей для генерації текстів чи зображень, таких як ChatGPT, MidJourney.

Корисні посилання

1. [мій курс по глибинному навчанню для аудіо](#)
2. [УКУ лабораторія з машинного навчання](#)
3. [deep learning book](#)