

A1. Sample index : 推導用 1...M、程式用 0...M-1 的差異？

是否有研究？

沒有專門研究，主要是 符號慣例差異。

- 數學：索引從 1 開始
- 程式：記憶體從 0 起算，因此以 **0-based index** 為標準

口語化版本

- 數學與程式的索引方式不同，但不影響 MSE 的自由度、估計量或梯度推導。
- 這只是「符號不同」，不是「統計或數學上的自由度問題」。

B1. 非 Gaussian (heavy-tailed、skewed) 誤差下，t-test / χ^2 CI 是否會失效？

是否要用 bootstrap 或 sandwich estimator？

已有大量研究

當誤差重尾或偏態， t/χ^2 這些 依賴常態假設 的信賴區間會出現 coverage failure。

解決方向

- Sandwich estimator (heteroskedastic-robust)
- Nonparametric bootstrap
- Bayesian posterior intervals

口語化版

- 如果資料的誤差「不乖」、不是常態，那你用教科書的 t-test 信賴區間可能會亂掉。
- 這種情況就該用更 robust 的做法，例如 bootstrap 或 sandwich estimator。

C1. 梯度爆炸／消失與權重矩陣 ill-conditioning

Normalization 與 Residual Connection 是否能穩定梯度？

有研究

這是深度學習最核心的數值穩定性問題之一。

參考文獻

He et al. (2016). *Deep Residual Learning for Image Recognition.*
<https://arxiv.org/abs/1512.03385>

口語化

- 殘差結構與 normalization 是讓梯度不要爆掉或消失的「安全閥」。

C2. tanh 換成 ReLU 或 sigmoid，還會有類似結果嗎？

有研究，但答案是「不完全」

不同 activation → 不同光滑性、梯度行為、收斂速度、表示能力。

參考文獻

Glorot et al. (2011). *Deep Sparse Rectifier Neural Networks.*
<https://proceedings.mlr.press/v15/glorot11a>

口語化

- 換 activation 會改變梯度性質，不能假設結果都一樣。

D1. Multiclass Softmax Loss 是否仍是凸的？

在 linear model 下是凸的

但在 deep network (非線性) 中就不是凸的。

參考文獻

Boyd & Vandenberghe. *Convex Optimization.*
<https://web.stanford.edu/~boyd/cvxbook/>

口語化

- Logistic regression 是凸的，但 deep softmax 不是。

D2. payoff function 有 kink ($\max(S-K, 0)$) ，SGD 要怎麼處理？

是否用 subgradient ?

有理論 (凸分析)

kink 處可使用 subgradient method。

參考文獻

Rockafellar. *Convex Analysis*.

<https://press.princeton.edu/books/paperback/9780691015866/convex-analysis>

口語化

- 看起來不平滑的 payoff (例如 max) 沒問題，只要用 subgradient 就能跑 SGD。

E. 判別分析 (LDA / GDA / QDA)

E1. 什麼情況應使用 LDA？什麼情況應使用 GDA？

有研究

- LDA : $\Sigma_0 = \Sigma_1$ ，邊界線性
- GDA : $\Sigma_0 \neq \Sigma_1$ ，邊界非線性

參考文獻

Hastie, Tibshirani, Friedman, *The Elements of Statistical Learning*.

<https://hastie.su.domains/ElemStatLearn/>

口語化

- 共變異數相同 \rightarrow LDA
- 共變異數不同 \rightarrow GDA/QDA。

E2. 小樣本 covariance 不準 \rightarrow LDA/GDA 是否失效？

需要 regularization 嗎？

有研究

可使用 shrinkage covariance (例如 Ledoit–Wolf)。

參考文獻

Ledoit & Wolf (2004). *A well-conditioned estimator for large-dimensional covariance matrices*.

[https://doi.org/10.1016/S0047-259X\(03\)00096-4](https://doi.org/10.1016/S0047-259X(03)00096-4)

口語化

- 樣本太少時，估計 covariance 很容易失真，需要 shrinkage。

E3. 為什麼 $\Sigma_0 \neq \Sigma_1$ 時邊界變二次曲面？

理論已完備

Posterior 的 log-likelihood 差值形成 quadratic form → QDA。

參考文獻

同上 ESL Ch. 4.

F1. SSM approximation 在高維流形上行為如何？

屬 open problem

目前沒有一致結論。

相關文獻 (但未解決此問題)

Gu et al. (2022). *Structured State Spaces*.

<https://arxiv.org/abs/2111.00396>

口語化

- SSM 在高維彎曲空間 (像人臉 manifold) 是否能維持穩定？
- 目前沒人完全知道。

F2. Ito's lemma 能否在流形／非歐幾里得空間保持同樣形式？

有研究 (Stochastic calculus on manifolds)

但會多出 Christoffel symbols，形式並不相同。

參考文獻

Hsu (2002). *Stochastic Analysis on Manifolds*.

<https://link.springer.com/book/10.1007/978-1-4612-0039-0>

口語化

- 在彎曲空間使用 Ito's lemma，需要多一些描述彎曲程度的修正項。

作業用gpt生成(按照功課指令使用)