جامعة الأخوين
ⵜⴰⵙⴷⴰⵡⵉⵜ ⵏ ⵉⵅⴰⵡⴰⵢⵏ
AL AKHAWAYN UNIVERSITY

Critique Notes: Claude Transcript Analysis
Project 2: AI-Assisted Weight Initialization
Mohamed Ali Kabiri

What Claude Got Right

1. Orthogonal Initialization:
   - Said orthogonal matrices keep lengths the same: ||Wx|| = ||x||
   - Said all singular values are 1
   - Said condition number is 1
   - Gave correct method: use QR decomposition

2. Gaussian Initialization:
   - Gave Xavier formula: $\sigma^2 = 2/(d_{in} + d_{out})$
   - Gave He formula: $\sigma^2 = 2/d_{in}$
   - Explained the math: to keep variance, need $d \cdot \sigma^2 = 1$

3. Uniform Initialization:
   - Gave correct formula: a = sqrt(3/d)
   - Said variance = $a^2/3$
   - Explained uniform matches Gaussian variance

What Claude Got Wrong

1. Big Mistake: σ = 1/d:
   - Claude said σ = 1/d would preserve variance
   - Our tests showed this is completely wrong
   - With σ = 1/d, variance drops to 0%
   - Gradients vanish completely

2. Confusing Recommendations:
   - Gave three different σ values without saying which to use
   - σ = 1/d (0.01), σ = 1/sqrt(d) (0.1), σ = sqrt(2/d) (0.1414)
   - No clear guidance on which is best

3. Small Errors:
   - Some math formulas were written poorly
   - Said to scale orthogonal matrices by gain g, but if $g \neq 1$, they're not orthogonal anymore
   - Some typos in the equations

What Claude Was Partly Right About

1. Uniform initialization stability:
   - Said bounded range [-a,a] is more stable
   - Our tests showed uniform works similarly to Gaussian $\sigma=1/\sqrt{d}$
   - But not necessarily "more stable" in our linear network

2. All methods aim for stability:
- True, but they achieve it very differently
- Orthogonal: exact mathematical guarantee
- Gaussian/Uniform: statistical, only works on average

Our Test Results vs Claude's Claims

1. Orthogonal: Claude was 100% right
- Condition number = 1.000
- Singular values all 1.000
- Gradients preserved exactly

2. Gaussian σ=1/d: Claude was 100% wrong
- Claude said: preserves variance
- Our test: 0% preservation (So wrong)
- Gradients vanished to near zero

3. Gaussian σ=1/√d: Claude was right
- Claude said: preserves variance
- Our test: 100.4% preservation

4. Uniform: Claude was mostly right
- Claude said: matches Gaussian variance
- Our test: 97.7% preservation
- Performs similarly to Gaussian σ=1/√d

Overall Assessment

So Claude was almost right except for one major mistake

The main problem is Claude said σ = 1/d would work, but it fails completely. This could mislead someone into using a bad initialization.

We can say as a conclusion that we need to always test AI math advice. The AI can give good starting points but can make serious mistakes that need checking.