Teacher Forcing

Training technique for sequence generation

Teacher Forcing

Uses **ground truth** as decoder input (not predictions)

Normal Inference

Uses **model's own predictions** as next input

How It Works

1. **During Training:** Model ignores its own predictions and always uses ground truth as next input
2. **Example:** Translating "I love cats" → Even if model incorrectly predicts "나는" → Still provide correct "나는" as next input
3. **During Inference:** No ground truth available, so model uses its own predictions as next input → Initial errors can accumulate

Calculation Example: Sequence Generation (3D Input → 1D Output)

Scenario: Input vector $[x_1, x_2, x_3]$ → Generate number sequence (e.g., [2, 5, 3])

Teacher Forcing (Training)

t=1: Input: [1.2, 0.8, 2.1] → Model predicts: **4** → ✋ **Ignore!** → Use ground truth **2** as next input

t=2: Input: [2, 0.8, 2.1] → Model predicts: **6** → ✋ **Ignore!**

Normal Inference (Testing)

t=1: Input: [1.2, 0.8, 2.1] → Model predicts: **4** → ⚠ **Use it!** → Use **4** as next input

t=2: Input: [4, 0.8, 2.1] → Model predicts: **7** → ⚠ **Use it!**

→ Use ground truth **5** as next input

t=3: Input: [5, 0.8, 2.1] → Model predicts: **4** → ✋ **Ignore!**
→ Use ground truth **3** as next input

✅ Always learns with correct context → Fast convergence

→ Use **7** as next input

t=3: Input: [7, 0.8, 2.1] → Model predicts: **9** → ⚠️ **Use it!**
→ Output complete

⚠️ Initial errors accumulate → Diverges from ground truth [2,5,3] (Exposure Bias)

✓ Benefits

- Accelerates training convergence
- Faster learning

⚠ Problem

- **Exposure bias:** train/test mismatch
- Lower inference quality

Solution: Scheduled Sampling

Gradually use predictions → Trade-off between training speed and inference quality

Scheduled Sampling Implementation

- **Early stage:** 100% ground truth (Pure Teacher Forcing)
- **Gradual transition:** Use ground truth with probability p, model predictions with (1-p) (p decreases over time)
- **Late stage:** Mostly use model predictions → Similar to actual inference environment