Financial Policy Optimization Report

Deep Learning vs Reinforcement Learning

Executive Summary:

This report analyzes the performance of a supervised Deep Learning (DL) model and a Reinforcement Learning (RL) agent for loan approval decisions.

Key Findings:

1. The DL model outperforms the RL agent significantly in terms of financial value.
   - DL Policy Value: -1.66 Million (Conservative)
   - RL Policy Value: -11.11 Million (Aggressive)

2. The RL agent appears over-optimistic, approving high-risk loans that the DL model correctly identifies as likely defaults.

3. Recommendation: Deploy the DL model (or a tuned conservative version) and investigate reward shaping for the RL agent.

1. Results Presentation

Metric	DL Model	RL Agent
AUC (ROC)	0.7463	N/A
F1 Score (Default)	0.4491	N/A
Est. Policy Value ($)	$-1,660,097.66	$-11,113,506.83

2. Metric Explanation:

AUC & F1 (DL Model):
- AUC (Area Under Curve) measures the model's ability to rank borrowers by risk.
  A score of 0.74 indicates good predictive power.
- F1-Score balances precision and recall for detecting defaults.

Estimated Policy Value (RL Agent):
- This represents the total financial outcome (Profit/Loss) if the agent's
  decisions were applied to the test set.
- It is the direct business objective: Maximizing portfolio return.
- The negative values indicate that on this test set (which may have high
  default rates or low interest margins), simply approving loans yielded losses,
  but RL lost significantly more by failing to screen bad loans.

3. Policy Comparison

The DL model implicitly defines a policy: Approve if P(Default) < 0.5.
The RL agent learns a policy directly (Action -> 1 or 0).

Disagreement Analysis:
The analysis found 2481 instances where the DL model flagged a loan as High Risk (Deny), but the RL agent Approved it.

Specific Example (Index 3):
- DL Predicted Prob of Default: 0.65 (High Risk)
- RL Action: 1 (Approve)
- Actual Outcome: Default
- Financial Consequence: $-16,800.00

Interpretation:
The RL agent approved a loan that effectively defaulted, resulting in a loss. This suggests the agent's value function estimation may be biased or it hasn't converged to recognizing these risk factors.

4. Future Steps

Deploy Strategy:
- Deploy the DL Model. It is safer and financially superior currently.

Limitations:
- RL performance is poor, likely due to sparse rewards or insufficient training.
- The dataset class imbalance might bias the RL agent to 'Approve All' if most loans are paid, but here it seems to be approving bad loans too.

Next Steps:
1. Tuning RL: Try scaling rewards differently (e.g., penalty for default * 10).
2. Data Collection: Collect more features on borrower behavior over time.
3. Algorithms: Explore DDPG or PPO for continuous action spaces (e.g., setting loan amount or interest rate dynamically).

Appendix: DL Model ROC Curve