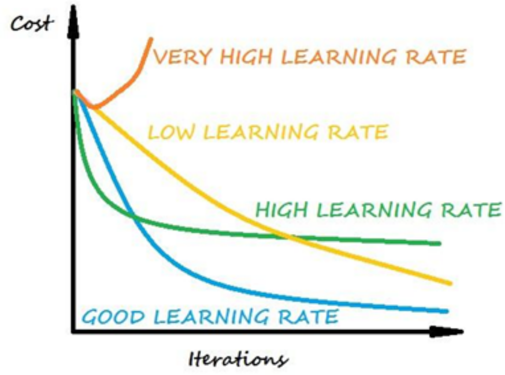
MAB Search in Supervised Learning: Adaptive Learning Rate Optimization

M. K. Guruprasad

Indian Institute of Information Technology, Design & Manufacturing, Kancheepuram
Mentor: Dr. Syed Shahul Hameed A S

Problem Statement

- Gradient Descent is a widely used cost optimization technique, and its performance heavily depends on the learning rate (LR).



- Very small LR $\rightarrow$ slow convergence
- Very large LR $\rightarrow$ divergence

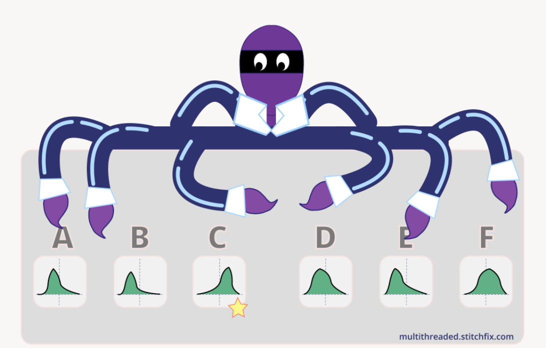
- We propose automating LR selection using Reinforcement Learning.

Objective

- We use a Reinforcement Learning technique called Multi-Armed Bandit (MAB).
- Automate learning rate selection during training.
- Apply MAB-based LR selection to a regression task.
- Compare with fixed learning rate strategies.

Multi-Armed Bandit (MAB)

- MAB is a Reinforcement Learning model based on slot machines.
- Arms yield rewards; the goal is to maximize long-term reward.
- Balances exploration vs. exploitation using ϵ-greedy:
  - Explore (random arm) with probability ϵ
  - Exploit (best arm) with probability $1 - \epsilon$



Applying MAB to Gradient Descent

- Arms: Learning rate candidates $\{0.1, 0.0001\}$
- Reward function:

$$R_i \leftarrow (1-\beta) \cdot R_i + \beta \cdot \mathrm{MSE}_t$$

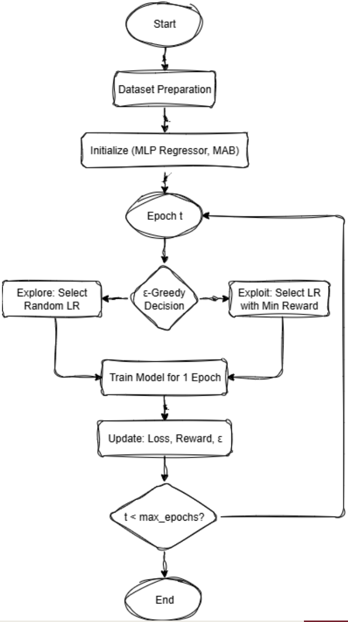
- Epsilon update (exploration decay):

$$\epsilon_t = \epsilon_{\min} + (\epsilon_{\max} - \epsilon_{\min}) \cdot e^{-\lambda t}$$

Data, Model & Training Setup

- **Dataset & Preprocessing:** California Housing, with normalization.
- **Model:** MLPRegressor configured as Linear Regressor:
  - No hidden layers (linear model)
  - Activation: Identity
  - Solver: SGD
- **Learning Rates:**
  - Fixed: $\{0.1, 0.01, 0.001, 0.0001\}$
  - MAB Actions: $\{0.1, 0.0001\}$
- **Evaluation Metric:**

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

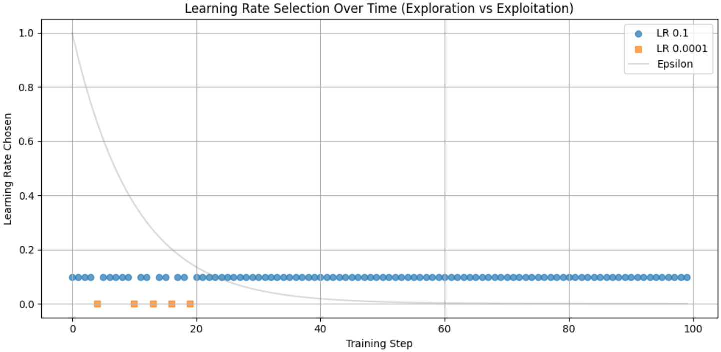


Results: Fixed vs. MAB Search

Learning Rate Strategy	Train MSE	Test MSE
0.1	0.5190	0.5630
0.0001	6.4710	5.8914
MAB Search	0.5287	0.5518

- MAB performs similar or slightly better than best fixed LR.
- Avoids poor performance from suboptimal LR.

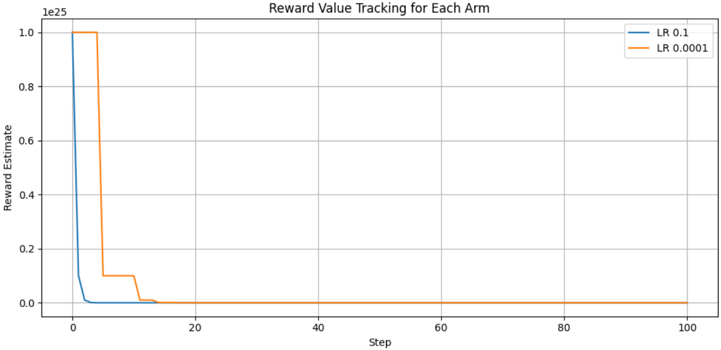
Learning Behavior – LR Selection & Epsilon Decay



Learning Rate Selection Over Time

- The learning rate gradually converges to the optimal value (0.1).
- Epsilon decay reduces exploration over time, encouraging exploitation of the best action.
- Shows effective adaptation of MAB to the training dynamics.

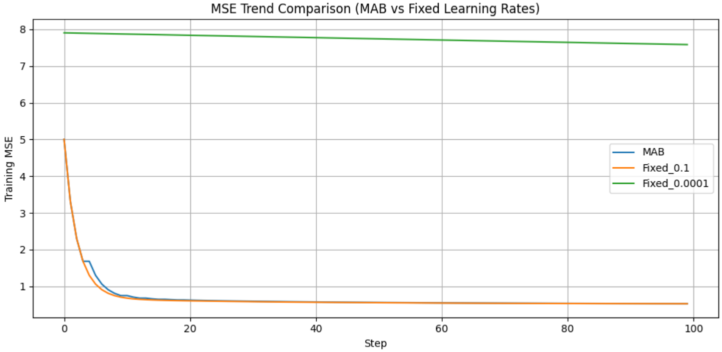
Learning Behavior – Reward Curve



Smoothed Reward Curve Over Time

- The reward (based on training MSE) decreases consistently.
- Reflects the MAB agent learning to avoid poor LR choices.
- Smoother reward indicates stability in performance over time.

Performance Comparison – MSE Trend



Train vs Test MSE Across LRs and MAB

- Fixed LR $= 0.0001$ underperforms due to underfitting.
- LR $= 0.1$ achieves lowest MSE on both train and test.
- MAB strategy closely tracks the best-performing LR, with slight variations during exploration.

Conclusion & Future Work

- Successfully applied MAB to automate LR selection in supervised learning.
- Model dynamically selects best LR $\rightarrow$ competitive performance.

Future directions:

- Apply to non-linear models and deeper networks
- Use larger or more complex datasets
- Experiment with different reward/epsilon decay strategies

Thank you!

Questions?