# Maman 11 By Guy Vitelson

## If you run this within Google Collab, Dont Worry!

**all the missing python files/directories/modules will be automatically feteched from my github repository**

**My GitHub Profile :** https://github.com/v1t3ls0n

**The Repository:** https://github.com/v1t3ls0n/ml_intro_course_mmn11

# Overview

## MNIST Digit Classification Using Perceptron Learning Algorithm (PLA)

**Objective:**
This notebook compares the performance of two variants of the Perceptron Learning Algorithm (PLA) on the MNIST digit classification task:

- **Clean PLA:** Standard perceptron without enhancements.
- **Pocket PLA:** Enhanced perceptron that stores the best-performing weights during training (using the Pocket algorithm).

**Dataset:**

- MNIST dataset consisting of 60,000 training samples and 10,000 test samples.
- The images are normalized to the range [0, 1] and a bias term is added, resulting in input samples with 785 features.

**Evaluation Metrics:**

- **Confusion Matrices:** Provides a detailed view of how well each digit is classified.
- **Overall Accuracy (ACC):** Defined as (\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN}).
- **Sensitivity (True Positive Rate, TPR):** For each digit, calculated as (\text{TPR} = \frac{TP}{TP + FN}), showing the model's ability to correctly identify the digit.
- **Selectivity (Specificity, TNR):** For each digit, calculated as (\text{TNR} = \frac{TN}{TN + FP}), showing the model's ability to correctly identify negatives.
- **Training and Testing Error Curves:** Visualized as a function of iteration for detailed analysis of learning dynamics.
- **Runtime:** The time taken to train the models.

**Goals:**

- Evaluate and compare the model accuracy and robustness between Clean PLA and Pocket PLA.
- Analyze and visualize the performance through confusion matrices, error curves, and summary plots (accuracy, sensitivity, selectivity, and runtime vs. the number of iterations).
- Provide a comprehensive discussion on how training iterations affect the decision boundaries and the overall performance, particularly in the one-vs-all classification setup.

This notebook integrates detailed quantitative evaluation with comprehensive visualizations to thoroughly analyze the multi-class Perceptron performance on the MNIST dataset.

# Imports

## External Code Imports (pip packages)

In [1]:

```
import os
import shutil
import sys
import logging
import numpy as np # type: ignore
import matplotlib.pyplot as plt # type: ignore
import seaborn as sns # type: ignore
import time
import pandas as pd
```

## Fetch Missing Files For Google Colab Env

In [2]:

```
# %%capture run_output
# %matplotlib inline

if sys.platform != 'win32': # check if we are running on google collab
  repo_url = "https://github.com/v1t3ls0n/ml_intro_course_mmn11"
  repo_name = "ml_intro_course_mmn11"
  from tqdm.notebook import tqdm # type: ignore


  # Clone the repository if it doesn't exist
  if not os.path.exists(repo_name):
    os.system(f"git clone {repo_url}")

  # Construct the path to the repository directory
  repo_path = os.path.join(os.getcwd(), repo_name)

  # Add the repository directory to the Python path
  if repo_path not in sys.path:
    sys.path.insert(0, repo_path)

  # --- Extract 'core' and 'notebooks' directories ---
  def extract_directories(source_dir, destination_dir, dir_names):
      for dir_name in dir_names:
          source_path = os.path.join(source_dir, dir_name)
          destination_path = os.path.join(destination_dir, dir_name)
          if os.path.exists(source_path):
              shutil.copytree(source_path, destination_path, dirs_exist_ok=True)

  destination_path = "."
  # Extract the directories
  extract_directories(repo_path, destination_path, ["core"])
  project_root = os.path.abspath(os.path.join(os.getcwd(), '..'))
  sys.path.insert(0, project_root)
  if os.path.exists("ml_intro_course_mmn11"):
    shutil.rmtree("ml_intro_course_mmn11")
  if os.path.exists("sample_data"):
    shutil.rmtree("sample_data")
else:
  from tqdm import tqdm  # type: ignore
  current_dir = os.getcwd()  # Current working directory
  project_root = os.path.abspath(os.path.join(current_dir, '..'))  # Root directory of t
he project
  sys.path.insert(0, project_root)
```

## Internal Code Imports (original code)

In [3]:

```
# ---------- Internal Code Imports ----------
```

```
# ========== Internal Code Imports ==========

#Logger
from core.logger.config import logger

# Data Preprocessing
from core.data.mnist_loader import load_mnist
from core.data.data_preprocessing import preprocess_data

# Models
from core.models.perceptron.multi_class_perceptron import MultiClassPerceptron
from core.models.logistic_regression.softmax_lregression import SoftmaxRegression
from core.models.linear_regression.linear_regression import  LinearRegression

# Performance & Plotting
from core.analysis.evaluation_functions import (
    evaluate_model,
    aggregate_iteration_losses,
    aggregate_iteration_losses_softmax
)

from core.analysis.plotting import (
    plot_confusion_matrix_annotated,
    plot_error_curves,
    plot_accuracy_vs_max_iter,
    plot_runtime_vs_max_iter,
    plot_performance_summary_extended,
    plot_train_curves_three_models,
    plot_metric_vs_learning_rate,
    plot_accuracy_vs_max_iter_4models,
    plot_runtime_vs_max_iter_4models,
    plot_accuracy_vs_runtime,
    plot_performance_summary_extended_by_runtime,
    plot_performance_summary_4models_by_runtime,
    plot_accuracy_vs_runtime_4models
)

logger = logging.getLogger("MyGlobalLogger") # configured in core/logger/config.py
```

# Choose Run Parameters (Significant Effect On Model's Runtime!)

In [4]:

```
######################################################################
# SEPARATE RUN PARAMETERS FOR PERCEPTRONS vs. REGRESSIONS
######################################################################

# Perceptrons (Clean & Pocket) iteration-based run
perceptron_max_iter_values = [20,50,100,1000]  # for Clean PLA & Pocket PLA
# Logging the run parameters
logger.info(f"=== Perceptron Run Parameters ===")
logger.info(f"max_iter_values = {perceptron_max_iter_values}")


# Regression (Softmax & Linear) run parameters.
learning_rates = [0.1]
iteration_counts = [100,1000,10000]
regression_run_configs = [
    {
        "label": f"LR={lr}/Iter={it}",
        "learning_rate": lr,
        "max_iter": it
    }
    for lr in learning_rates
    for it in iteration_counts
]

logger.info(f"=== Regression Run Parameters ===")
```

```
for cfg in regression_run_configs:
    logger.info(f"{cfg['label']} -> learning_rate={cfg['learning_rate']}, max_iter={cfg[
'max_iter']}")
```

```
2025-03-18 21:25:45,596 - INFO - === Perceptron Run Parameters ===
2025-03-18 21:25:45,597 - INFO - max_iter_values = [20, 50, 100, 1000]
2025-03-18 21:25:45,597 - INFO - === Regression Run Parameters ===
2025-03-18 21:25:45,598 - INFO - LR=0.1/Iter=100 -> learning_rate=0.1, max_iter=100
2025-03-18 21:25:45,598 - INFO - LR=0.1/Iter=1000 -> learning_rate=0.1, max_iter=1000
2025-03-18 21:25:45,599 - INFO - LR=0.1/Iter=10000 -> learning_rate=0.1, max_iter=10000
2025-03-18 21:25:45,597 - INFO - max_iter_values = [20, 50, 100, 1000]
2025-03-18 21:25:45,597 - INFO - === Regression Run Parameters ===
2025-03-18 21:25:45,598 - INFO - LR=0.1/Iter=100 -> learning_rate=0.1, max_iter=100
2025-03-18 21:25:45,598 - INFO - LR=0.1/Iter=1000 -> learning_rate=0.1, max_iter=1000
2025-03-18 21:25:45,599 - INFO - LR=0.1/Iter=10000 -> learning_rate=0.1, max_iter=10000
```

# Load and Preprocess the MNIST Dataset

In [5]:

```
'''
We'll load the MNIST dataset using our custom loader (`mnist_loader`) and then apply prep
rocessing (`data_preprocessing`).
The preprocessing step normalizes each image to the range [0, 1] and adds a bias term, re
sulting in input samples with 785 features.
This setup ensures that the training set contains 60,000 samples and the test set 10,000
samples, preparing the data for the subsequent classification tasks.
'''

# New section
# Load raw MNIST data (X: images, y: labels)
X_raw, y_raw = load_mnist()


logger.info("Raw MNIST data shapes: X_raw: %s, y_raw: %s", X_raw.shape, y_raw.shape)

# Preprocess (normalize & add bias = True)
X = preprocess_data(X_raw, add_bias=True, normalize=True)
logger.info("Preprocessed shape: %s", X.shape)

# Split into train/test manually or with 60k/10k as the task suggests
X_train, y_train = X[:60000], y_raw[:60000]
X_test,  y_test  = X[60000:], y_raw[60000:]

logger.info("Train set: X_train: %s, y_train: %s", X_train.shape, y_train.shape)
logger.info("Test set: X_test: %s, y_test: %s", X_test.shape, y_test.shape)
```

```
2025-03-18 21:25:47,833 - INFO - Raw MNIST data shapes: X_raw: (70000, 784), y_raw: (7000
0,)
2025-03-18 21:25:48,000 - INFO - Preprocessed shape: (70000, 785)
2025-03-18 21:25:48,002 - INFO - Train set: X_train: (60000, 785), y_train: (60000,)
2025-03-18 21:25:48,002 - INFO - Test set: X_test: (10000, 785), y_test: (10000,)
```

# Train

In [6]:

```
# ============================================================
# TRAINING CELL
# ============================================================

# 1) Dictionaries to store trained models
trained_models_clean   = {}
trained_models_pocket  = {}
trained_models_softmax = {}
trained_models_linear  = {}

# 2) Train Regression Models (Softmax & Linear)
```

```
logger.info("=== TRAINING REGRESSION MODELS (Softmax & Linear) ===")
for cfg in tqdm(regression_run_configs, desc="Train Regressions"):
    lr_val = cfg["learning_rate"]
    max_iter_val = cfg["max_iter"]
    label = cfg["label"]  # e.g. "LR=0.001/Iter=1000"

    # --- Softmax ---
    logger.info(f"--- Softmax {label} ---")
    s_model = SoftmaxRegression(
        num_classes=10,
        max_iter=max_iter_val,
        learning_rate=lr_val,
        adaptive_lr=True
    )
    s_model.fit(X_train, y_train)
    trained_models_softmax[(lr_val, max_iter_val)] = s_model

    # --- Linear ---
    logger.info(f"--- Linear Regression {label} ---")
    lin_model = LinearRegression(
        num_classes=10,
        max_iter=max_iter_val,
        learning_rate=lr_val,
        adaptive_lr=True,
        early_stopping=False
    )
    lin_model.fit(X_train, y_train)
    trained_models_linear[(lr_val, max_iter_val)] = lin_model

logger.info("Training complete for Softmax and Linear.")

# 3) Train Perceptron Models (Clean & Pocket)
logger.info("=== TRAINING PERCEPTRON MODELS (Clean & Pocket) ===")
for max_iter in tqdm(perceptron_max_iter_values, desc="Train Clean & Pocket"):
    logger.info(f"--- Clean PLA, max_iter={max_iter} ---")
    clean_perc = MultiClassPerceptron(num_classes=10, max_iter=max_iter, use_pocket=Fals
e)
    clean_perc.fit(X_train, y_train)
    trained_models_clean[max_iter] = clean_perc

    logger.info(f"--- Pocket PLA, max_iter={max_iter} ---")
    pocket_perc = MultiClassPerceptron(num_classes=10, max_iter=max_iter, use_pocket=Tru
e)
    pocket_perc.fit(X_train, y_train)
    trained_models_pocket[max_iter] = pocket_perc

logger.info("Training complete for Clean PLA and Pocket PLA.")
logger.info("=== ALL TRAINING COMPLETE ===")
```

```
2025-03-18 21:25:48,008 - INFO - === TRAINING REGRESSION MODELS (Softmax & Linear) ===
Train Regressions:   0%|          | 0/3 [00:00<?, ?it/s]2025-03-18 21:25:48,010 - INFO -
--- Softmax LR=0.1/Iter=100 ---
Train Regressions:   0%|          | 0/3 [00:00<?, ?it/s]2025-03-18 21:25:48,010 - INFO -
--- Softmax LR=0.1/Iter=100 ---
2025-03-18 21:25:48,054 - INFO - Iter 1/100, Loss: 2.4006, Avg Adaptive LR: 14.067368
2025-03-18 21:25:48,480 - INFO - Iter 11/100, Loss: 0.4338, Avg Adaptive LR: 3.044223
2025-03-18 21:25:48,888 - INFO - Iter 21/100, Loss: 0.3772, Avg Adaptive LR: 3.040986
2025-03-18 21:25:49,299 - INFO - Iter 31/100, Loss: 0.3520, Avg Adaptive LR: 3.039560
2025-03-18 21:25:49,720 - INFO - Iter 41/100, Loss: 0.3361, Avg Adaptive LR: 3.038652
2025-03-18 21:25:50,131 - INFO - Iter 51/100, Loss: 0.3248, Avg Adaptive LR: 3.038002
2025-03-18 21:25:50,543 - INFO - Iter 61/100, Loss: 0.3162, Avg Adaptive LR: 3.037505
2025-03-18 21:25:50,944 - INFO - Iter 71/100, Loss: 0.3093, Avg Adaptive LR: 3.037109
2025-03-18 21:25:51,356 - INFO - Iter 81/100, Loss: 0.3037, Avg Adaptive LR: 3.036782
2025-03-18 21:25:51,756 - INFO - Iter 91/100, Loss: 0.2989, Avg Adaptive LR: 3.036506
2025-03-18 21:25:52,119 - INFO - SoftmaxRegression training completed in 4.11 seconds.
2025-03-18 21:25:52,120 - INFO - --- Linear Regression LR=0.1/Iter=100 ---
2025-03-18 21:25:55,318 - INFO - Iter 100/100, Loss: 0.7147, Gradient Norm: 16.0981, Avg
Adaptive LR: 1.3971574443060855
2025-03-18 21:25:55,318 - INFO - LinearRegressionClassifier training completed in 3.20 se
conds.
Train Regressions:  33%|███       | 1/3 [00:07<00:14,  7.31s/it]2025-03-18 21:25:55,319 -
INFO - --- Softmax LR=0.1/Iter=1000 ---
```

```
2025-03-18 21:25:55,358 - INFO - Iter 1/1000, Loss: 2.3143, Avg Adaptive LR: 13.534203
2025-03-18 21:25:55,760 - INFO - Iter 11/1000, Loss: 0.4351, Avg Adaptive LR: 3.351356
2025-03-18 21:25:56,174 - INFO - Iter 21/1000, Loss: 0.3740, Avg Adaptive LR: 3.346282
2025-03-18 21:25:56,583 - INFO - Iter 31/1000, Loss: 0.3483, Avg Adaptive LR: 3.344529
2025-03-18 21:25:56,986 - INFO - Iter 41/1000, Loss: 0.3324, Avg Adaptive LR: 3.343431
2025-03-18 21:25:57,392 - INFO - Iter 51/1000, Loss: 0.3212, Avg Adaptive LR: 3.342651
2025-03-18 21:25:57,794 - INFO - Iter 61/1000, Loss: 0.3128, Avg Adaptive LR: 3.342059
2025-03-18 21:25:58,195 - INFO - Iter 71/1000, Loss: 0.3060, Avg Adaptive LR: 3.341587
2025-03-18 21:25:58,608 - INFO - Iter 81/1000, Loss: 0.3005, Avg Adaptive LR: 3.341200
2025-03-18 21:25:59,008 - INFO - Iter 91/1000, Loss: 0.2959, Avg Adaptive LR: 3.340875
2025-03-18 21:25:59,419 - INFO - Iter 101/1000, Loss: 0.2919, Avg Adaptive LR: 3.340596
2025-03-18 21:25:59,829 - INFO - Iter 111/1000, Loss: 0.2885, Avg Adaptive LR: 3.340353
2025-03-18 21:26:00,235 - INFO - Iter 121/1000, Loss: 0.2854, Avg Adaptive LR: 3.340139
2025-03-18 21:26:00,641 - INFO - Iter 131/1000, Loss: 0.2827, Avg Adaptive LR: 3.339949
2025-03-18 21:26:01,042 - INFO - Iter 141/1000, Loss: 0.2803, Avg Adaptive LR: 3.339778
2025-03-18 21:26:01,445 - INFO - Iter 151/1000, Loss: 0.2781, Avg Adaptive LR: 3.339624
2025-03-18 21:26:01,847 - INFO - Iter 161/1000, Loss: 0.2761, Avg Adaptive LR: 3.339483
2025-03-18 21:26:02,256 - INFO - Iter 171/1000, Loss: 0.2743, Avg Adaptive LR: 3.339354
2025-03-18 21:26:02,656 - INFO - Iter 181/1000, Loss: 0.2726, Avg Adaptive LR: 3.339235
2025-03-18 21:26:03,071 - INFO - Iter 191/1000, Loss: 0.2710, Avg Adaptive LR: 3.339125
2025-03-18 21:26:03,478 - INFO - Iter 201/1000, Loss: 0.2696, Avg Adaptive LR: 3.339023
2025-03-18 21:26:03,885 - INFO - Iter 211/1000, Loss: 0.2682, Avg Adaptive LR: 3.338928
2025-03-18 21:26:04,291 - INFO - Iter 221/1000, Loss: 0.2670, Avg Adaptive LR: 3.338839
2025-03-18 21:26:04,688 - INFO - Iter 231/1000, Loss: 0.2658, Avg Adaptive LR: 3.338755
2025-03-18 21:26:05,085 - INFO - Iter 241/1000, Loss: 0.2647, Avg Adaptive LR: 3.338676
2025-03-18 21:26:05,497 - INFO - Iter 251/1000, Loss: 0.2636, Avg Adaptive LR: 3.338601
2025-03-18 21:26:05,893 - INFO - Iter 261/1000, Loss: 0.2626, Avg Adaptive LR: 3.338531
2025-03-18 21:26:06,297 - INFO - Iter 271/1000, Loss: 0.2617, Avg Adaptive LR: 3.338464
2025-03-18 21:26:06,694 - INFO - Iter 281/1000, Loss: 0.2608, Avg Adaptive LR: 3.338400
2025-03-18 21:26:07,094 - INFO - Iter 291/1000, Loss: 0.2599, Avg Adaptive LR: 3.338339
2025-03-18 21:26:07,499 - INFO - Iter 301/1000, Loss: 0.2591, Avg Adaptive LR: 3.338282
2025-03-18 21:26:07,898 - INFO - Iter 311/1000, Loss: 0.2583, Avg Adaptive LR: 3.338226
2025-03-18 21:26:08,303 - INFO - Iter 321/1000, Loss: 0.2576, Avg Adaptive LR: 3.338173
2025-03-18 21:26:08,701 - INFO - Iter 331/1000, Loss: 0.2569, Avg Adaptive LR: 3.338122
2025-03-18 21:26:09,100 - INFO - Iter 341/1000, Loss: 0.2562, Avg Adaptive LR: 3.338074
2025-03-18 21:26:09,502 - INFO - Iter 351/1000, Loss: 0.2555, Avg Adaptive LR: 3.338027
2025-03-18 21:26:09,896 - INFO - Iter 361/1000, Loss: 0.2549, Avg Adaptive LR: 3.337982
2025-03-18 21:26:10,308 - INFO - Iter 371/1000, Loss: 0.2543, Avg Adaptive LR: 3.337938
2025-03-18 21:26:10,705 - INFO - Iter 381/1000, Loss: 0.2537, Avg Adaptive LR: 3.337896
2025-03-18 21:26:11,107 - INFO - Iter 391/1000, Loss: 0.2531, Avg Adaptive LR: 3.337856
2025-03-18 21:26:11,515 - INFO - Iter 401/1000, Loss: 0.2525, Avg Adaptive LR: 3.337817
2025-03-18 21:26:11,915 - INFO - Iter 411/1000, Loss: 0.2520, Avg Adaptive LR: 3.337779
2025-03-18 21:26:12,329 - INFO - Iter 421/1000, Loss: 0.2515, Avg Adaptive LR: 3.337742
2025-03-18 21:26:12,727 - INFO - Iter 431/1000, Loss: 0.2510, Avg Adaptive LR: 3.337707
2025-03-18 21:26:13,124 - INFO - Iter 441/1000, Loss: 0.2505, Avg Adaptive LR: 3.337672
2025-03-18 21:26:13,544 - INFO - Iter 451/1000, Loss: 0.2500, Avg Adaptive LR: 3.337639
2025-03-18 21:26:13,956 - INFO - Iter 461/1000, Loss: 0.2496, Avg Adaptive LR: 3.337606
2025-03-18 21:26:14,359 - INFO - Iter 471/1000, Loss: 0.2491, Avg Adaptive LR: 3.337575
2025-03-18 21:26:14,761 - INFO - Iter 481/1000, Loss: 0.2487, Avg Adaptive LR: 3.337544
2025-03-18 21:26:15,158 - INFO - Iter 491/1000, Loss: 0.2483, Avg Adaptive LR: 3.337514
2025-03-18 21:26:15,562 - INFO - Iter 501/1000, Loss: 0.2479, Avg Adaptive LR: 3.337485
2025-03-18 21:26:15,963 - INFO - Iter 511/1000, Loss: 0.2475, Avg Adaptive LR: 3.337457
2025-03-18 21:26:16,368 - INFO - Iter 521/1000, Loss: 0.2471, Avg Adaptive LR: 3.337429
2025-03-18 21:26:16,762 - INFO - Iter 531/1000, Loss: 0.2467, Avg Adaptive LR: 3.337402
2025-03-18 21:26:17,155 - INFO - Iter 541/1000, Loss: 0.2463, Avg Adaptive LR: 3.337376
2025-03-18 21:26:17,562 - INFO - Iter 551/1000, Loss: 0.2459, Avg Adaptive LR: 3.337350
2025-03-18 21:26:17,956 - INFO - Iter 561/1000, Loss: 0.2456, Avg Adaptive LR: 3.337325
2025-03-18 21:26:18,373 - INFO - Iter 571/1000, Loss: 0.2452, Avg Adaptive LR: 3.337300
2025-03-18 21:26:18,771 - INFO - Iter 581/1000, Loss: 0.2449, Avg Adaptive LR: 3.337277
2025-03-18 21:26:19,176 - INFO - Iter 591/1000, Loss: 0.2446, Avg Adaptive LR: 3.337253
2025-03-18 21:26:19,582 - INFO - Iter 601/1000, Loss: 0.2443, Avg Adaptive LR: 3.337230
2025-03-18 21:26:19,983 - INFO - Iter 611/1000, Loss: 0.2439, Avg Adaptive LR: 3.337208
2025-03-18 21:26:20,388 - INFO - Iter 621/1000, Loss: 0.2436, Avg Adaptive LR: 3.337186
2025-03-18 21:26:20,791 - INFO - Iter 631/1000, Loss: 0.2433, Avg Adaptive LR: 3.337164
2025-03-18 21:26:21,189 - INFO - Iter 641/1000, Loss: 0.2430, Avg Adaptive LR: 3.337143
2025-03-18 21:26:21,600 - INFO - Iter 651/1000, Loss: 0.2427, Avg Adaptive LR: 3.337123
2025-03-18 21:26:21,999 - INFO - Iter 661/1000, Loss: 0.2425, Avg Adaptive LR: 3.337103
2025-03-18 21:26:22,404 - INFO - Iter 671/1000, Loss: 0.2422, Avg Adaptive LR: 3.337083
2025-03-18 21:26:22,801 - INFO - Iter 681/1000, Loss: 0.2419, Avg Adaptive LR: 3.337063
2025-03-18 21:26:23,199 - INFO - Iter 691/1000, Loss: 0.2416, Avg Adaptive LR: 3.337044
2025-03-18 21:26:23,619 - INFO - Iter 701/1000, Loss: 0.2414, Avg Adaptive LR: 3.337025
2025-03-18 21:26:24,024 - INFO - Iter 711/1000, Loss: 0.2411, Avg Adaptive LR: 3.337007
```

```
2025-03-18 21:26:24,440 - INFO - Iter 721/1000, Loss: 0.2409, Avg Adaptive LR: 3.336989
2025-03-18 21:26:24,842 - INFO - Iter 731/1000, Loss: 0.2406, Avg Adaptive LR: 3.336971
2025-03-18 21:26:25,251 - INFO - Iter 741/1000, Loss: 0.2404, Avg Adaptive LR: 3.336954
2025-03-18 21:26:25,647 - INFO - Iter 751/1000, Loss: 0.2401, Avg Adaptive LR: 3.336937
2025-03-18 21:26:26,042 - INFO - Iter 761/1000, Loss: 0.2399, Avg Adaptive LR: 3.336920
2025-03-18 21:26:26,444 - INFO - Iter 771/1000, Loss: 0.2396, Avg Adaptive LR: 3.336903
2025-03-18 21:26:26,848 - INFO - Iter 781/1000, Loss: 0.2394, Avg Adaptive LR: 3.336887
2025-03-18 21:26:27,254 - INFO - Iter 791/1000, Loss: 0.2392, Avg Adaptive LR: 3.336871
2025-03-18 21:26:27,655 - INFO - Iter 801/1000, Loss: 0.2390, Avg Adaptive LR: 3.336855
2025-03-18 21:26:28,056 - INFO - Iter 811/1000, Loss: 0.2387, Avg Adaptive LR: 3.336840
2025-03-18 21:26:28,466 - INFO - Iter 821/1000, Loss: 0.2385, Avg Adaptive LR: 3.336825
2025-03-18 21:26:28,862 - INFO - Iter 831/1000, Loss: 0.2383, Avg Adaptive LR: 3.336810
2025-03-18 21:26:29,272 - INFO - Iter 841/1000, Loss: 0.2381, Avg Adaptive LR: 3.336795
2025-03-18 21:26:29,675 - INFO - Iter 851/1000, Loss: 0.2379, Avg Adaptive LR: 3.336780
2025-03-18 21:26:30,080 - INFO - Iter 861/1000, Loss: 0.2377, Avg Adaptive LR: 3.336766
2025-03-18 21:26:30,486 - INFO - Iter 871/1000, Loss: 0.2375, Avg Adaptive LR: 3.336752
2025-03-18 21:26:30,899 - INFO - Iter 881/1000, Loss: 0.2373, Avg Adaptive LR: 3.336738
2025-03-18 21:26:31,304 - INFO - Iter 891/1000, Loss: 0.2371, Avg Adaptive LR: 3.336724
2025-03-18 21:26:31,696 - INFO - Iter 901/1000, Loss: 0.2369, Avg Adaptive LR: 3.336711
2025-03-18 21:26:32,092 - INFO - Iter 911/1000, Loss: 0.2367, Avg Adaptive LR: 3.336698
2025-03-18 21:26:32,496 - INFO - Iter 921/1000, Loss: 0.2366, Avg Adaptive LR: 3.336684
2025-03-18 21:26:32,892 - INFO - Iter 931/1000, Loss: 0.2364, Avg Adaptive LR: 3.336672
2025-03-18 21:26:33,295 - INFO - Iter 941/1000, Loss: 0.2362, Avg Adaptive LR: 3.336659
2025-03-18 21:26:33,695 - INFO - Iter 951/1000, Loss: 0.2360, Avg Adaptive LR: 3.336646
2025-03-18 21:26:34,098 - INFO - Iter 961/1000, Loss: 0.2358, Avg Adaptive LR: 3.336634
2025-03-18 21:26:34,519 - INFO - Iter 971/1000, Loss: 0.2357, Avg Adaptive LR: 3.336622
2025-03-18 21:26:34,916 - INFO - Iter 981/1000, Loss: 0.2355, Avg Adaptive LR: 3.336609
2025-03-18 21:26:35,330 - INFO - Iter 991/1000, Loss: 0.2353, Avg Adaptive LR: 3.336598
2025-03-18 21:26:35,690 - INFO - SoftmaxRegression training completed in 40.37 seconds.
2025-03-18 21:26:35,691 - INFO - --- Linear Regression LR=0.1/Iter=1000 ---
2025-03-18 21:26:38,822 - INFO - Iter 100/1000, Loss: 0.7348, Gradient Norm: 16.3445, Avg
Adaptive LR: 1.3971218005437425
2025-03-18 21:26:41,998 - INFO - Iter 200/1000, Loss: 0.4130, Gradient Norm: 11.9788, Avg
Adaptive LR: 0.9913774491410133
2025-03-18 21:26:45,342 - INFO - Iter 300/1000, Loss: 0.2897, Gradient Norm: 9.8017, Avg
Adaptive LR: 0.8108005591183135
2025-03-18 21:26:48,670 - INFO - Iter 400/1000, Loss: 0.2214, Gradient Norm: 8.3568, Avg
Adaptive LR: 0.702854553981635
2025-03-18 21:26:51,986 - INFO - Iter 500/1000, Loss: 0.1827, Gradient Norm: 7.4142, Avg
Adaptive LR: 0.6291135396210678
2025-03-18 21:26:55,332 - INFO - Iter 600/1000, Loss: 0.1514, Gradient Norm: 6.5548, Avg
Adaptive LR: 0.5745581695686357
2025-03-18 21:26:58,660 - INFO - Iter 700/1000, Loss: 0.1321, Gradient Norm: 5.9598, Avg
Adaptive LR: 0.5321583299184253
2025-03-18 21:27:01,980 - INFO - Iter 800/1000, Loss: 0.1166, Gradient Norm: 5.4377, Avg
Adaptive LR: 0.49793763751408765
2025-03-18 21:27:05,340 - INFO - Iter 900/1000, Loss: 0.1051, Gradient Norm: 5.0141, Avg
Adaptive LR: 0.46959128033932185
2025-03-18 21:27:08,637 - INFO - Iter 1000/1000, Loss: 0.0960, Gradient Norm: 4.6565, Avg
Adaptive LR: 0.44559773839234845
2025-03-18 21:27:08,637 - INFO - LinearRegressionClassifier training completed in 32.95 s
econds.
Train Regressions:  67%|██████    | 2/3 [01:20<00:46, 46.14s/it]2025-03-18 21:27:08,638 -
INFO - --- Softmax LR=0.1/Iter=10000 ---
2025-03-18 21:27:08,680 - INFO - Iter 1/10000, Loss: 2.3145, Avg Adaptive LR: 15.220312
2025-03-18 21:27:09,079 - INFO - Iter 11/10000, Loss: 0.4765, Avg Adaptive LR: 3.326323
2025-03-18 21:27:09,483 - INFO - Iter 21/10000, Loss: 0.3710, Avg Adaptive LR: 3.305129
2025-03-18 21:27:09,890 - INFO - Iter 31/10000, Loss: 0.3469, Avg Adaptive LR: 3.303520
2025-03-18 21:27:10,298 - INFO - Iter 41/10000, Loss: 0.3316, Avg Adaptive LR: 3.302486
2025-03-18 21:27:10,692 - INFO - Iter 51/10000, Loss: 0.3207, Avg Adaptive LR: 3.301743
2025-03-18 21:27:11,086 - INFO - Iter 61/10000, Loss: 0.3124, Avg Adaptive LR: 3.301174
2025-03-18 21:27:11,489 - INFO - Iter 71/10000, Loss: 0.3057, Avg Adaptive LR: 3.300720
2025-03-18 21:27:11,888 - INFO - Iter 81/10000, Loss: 0.3002, Avg Adaptive LR: 3.300346
2025-03-18 21:27:12,292 - INFO - Iter 91/10000, Loss: 0.2957, Avg Adaptive LR: 3.300031
2025-03-18 21:27:12,691 - INFO - Iter 101/10000, Loss: 0.2917, Avg Adaptive LR: 3.299761
2025-03-18 21:27:13,098 - INFO - Iter 111/10000, Loss: 0.2883, Avg Adaptive LR: 3.299526
2025-03-18 21:27:13,501 - INFO - Iter 121/10000, Loss: 0.2853, Avg Adaptive LR: 3.299318
2025-03-18 21:27:13,903 - INFO - Iter 131/10000, Loss: 0.2826, Avg Adaptive LR: 3.299134
2025-03-18 21:27:14,314 - INFO - Iter 141/10000, Loss: 0.2802, Avg Adaptive LR: 3.298968
2025-03-18 21:27:14,710 - INFO - Iter 151/10000, Loss: 0.2780, Avg Adaptive LR: 3.298817
2025-03-18 21:27:15,141 - INFO - Iter 161/10000, Loss: 0.2760, Avg Adaptive LR: 3.298680
2025-03-18 21:27:15,548 - INFO - Iter 171/10000, Loss: 0.2742, Avg Adaptive LR: 3.298554
```

```
2025-03-18 21:27:15,952 - INFO - Iter 181/10000, Loss: 0.2725, Avg Adaptive LR: 3.298438
2025-03-18 21:27:16,367 - INFO - Iter 191/10000, Loss: 0.2709, Avg Adaptive LR: 3.298331
2025-03-18 21:27:16,784 - INFO - Iter 201/10000, Loss: 0.2695, Avg Adaptive LR: 3.298231
2025-03-18 21:27:17,192 - INFO - Iter 211/10000, Loss: 0.2681, Avg Adaptive LR: 3.298138
2025-03-18 21:27:17,602 - INFO - Iter 221/10000, Loss: 0.2669, Avg Adaptive LR: 3.298050
2025-03-18 21:27:18,008 - INFO - Iter 231/10000, Loss: 0.2657, Avg Adaptive LR: 3.297968
2025-03-18 21:27:18,416 - INFO - Iter 241/10000, Loss: 0.2645, Avg Adaptive LR: 3.297891
2025-03-18 21:27:18,820 - INFO - Iter 251/10000, Loss: 0.2635, Avg Adaptive LR: 3.297818
2025-03-18 21:27:19,228 - INFO - Iter 261/10000, Loss: 0.2625, Avg Adaptive LR: 3.297749
2025-03-18 21:27:19,640 - INFO - Iter 271/10000, Loss: 0.2615, Avg Adaptive LR: 3.297683
2025-03-18 21:27:20,046 - INFO - Iter 281/10000, Loss: 0.2606, Avg Adaptive LR: 3.297620
2025-03-18 21:27:20,459 - INFO - Iter 291/10000, Loss: 0.2598, Avg Adaptive LR: 3.297561
2025-03-18 21:27:20,876 - INFO - Iter 301/10000, Loss: 0.2589, Avg Adaptive LR: 3.297504
2025-03-18 21:27:21,291 - INFO - Iter 311/10000, Loss: 0.2582, Avg Adaptive LR: 3.297450
2025-03-18 21:27:21,700 - INFO - Iter 321/10000, Loss: 0.2574, Avg Adaptive LR: 3.297398
2025-03-18 21:27:22,113 - INFO - Iter 331/10000, Loss: 0.2567, Avg Adaptive LR: 3.297348
2025-03-18 21:27:22,525 - INFO - Iter 341/10000, Loss: 0.2560, Avg Adaptive LR: 3.297300
2025-03-18 21:27:22,922 - INFO - Iter 351/10000, Loss: 0.2553, Avg Adaptive LR: 3.297254
2025-03-18 21:27:23,351 - INFO - Iter 361/10000, Loss: 0.2547, Avg Adaptive LR: 3.297210
2025-03-18 21:27:23,754 - INFO - Iter 371/10000, Loss: 0.2541, Avg Adaptive LR: 3.297167
2025-03-18 21:27:24,155 - INFO - Iter 381/10000, Loss: 0.2535, Avg Adaptive LR: 3.297126
2025-03-18 21:27:24,573 - INFO - Iter 391/10000, Loss: 0.2529, Avg Adaptive LR: 3.297086
2025-03-18 21:27:24,978 - INFO - Iter 401/10000, Loss: 0.2523, Avg Adaptive LR: 3.297048
2025-03-18 21:27:25,404 - INFO - Iter 411/10000, Loss: 0.2518, Avg Adaptive LR: 3.297011
2025-03-18 21:27:25,803 - INFO - Iter 421/10000, Loss: 0.2513, Avg Adaptive LR: 3.296975
2025-03-18 21:27:26,204 - INFO - Iter 431/10000, Loss: 0.2508, Avg Adaptive LR: 3.296940
2025-03-18 21:27:26,625 - INFO - Iter 441/10000, Loss: 0.2503, Avg Adaptive LR: 3.296906
2025-03-18 21:27:27,034 - INFO - Iter 451/10000, Loss: 0.2498, Avg Adaptive LR: 3.296873
2025-03-18 21:27:27,446 - INFO - Iter 461/10000, Loss: 0.2493, Avg Adaptive LR: 3.296842
2025-03-18 21:27:27,842 - INFO - Iter 471/10000, Loss: 0.2489, Avg Adaptive LR: 3.296811
2025-03-18 21:27:28,247 - INFO - Iter 481/10000, Loss: 0.2485, Avg Adaptive LR: 3.296781
2025-03-18 21:27:28,643 - INFO - Iter 491/10000, Loss: 0.2480, Avg Adaptive LR: 3.296751
2025-03-18 21:27:29,044 - INFO - Iter 501/10000, Loss: 0.2476, Avg Adaptive LR: 3.296723
2025-03-18 21:27:29,449 - INFO - Iter 511/10000, Loss: 0.2472, Avg Adaptive LR: 3.296695
2025-03-18 21:27:29,851 - INFO - Iter 521/10000, Loss: 0.2468, Avg Adaptive LR: 3.296668
2025-03-18 21:27:30,256 - INFO - Iter 531/10000, Loss: 0.2465, Avg Adaptive LR: 3.296642
2025-03-18 21:27:30,661 - INFO - Iter 541/10000, Loss: 0.2461, Avg Adaptive LR: 3.296616
2025-03-18 21:27:31,056 - INFO - Iter 551/10000, Loss: 0.2457, Avg Adaptive LR: 3.296591
2025-03-18 21:27:31,463 - INFO - Iter 561/10000, Loss: 0.2454, Avg Adaptive LR: 3.296567
2025-03-18 21:27:31,858 - INFO - Iter 571/10000, Loss: 0.2450, Avg Adaptive LR: 3.296543
2025-03-18 21:27:32,263 - INFO - Iter 581/10000, Loss: 0.2447, Avg Adaptive LR: 3.296519
2025-03-18 21:27:32,667 - INFO - Iter 591/10000, Loss: 0.2444, Avg Adaptive LR: 3.296497
2025-03-18 21:27:33,069 - INFO - Iter 601/10000, Loss: 0.2440, Avg Adaptive LR: 3.296474
2025-03-18 21:27:33,479 - INFO - Iter 611/10000, Loss: 0.2437, Avg Adaptive LR: 3.296452
2025-03-18 21:27:33,921 - INFO - Iter 621/10000, Loss: 0.2434, Avg Adaptive LR: 3.296431
2025-03-18 21:27:34,327 - INFO - Iter 631/10000, Loss: 0.2431, Avg Adaptive LR: 3.296410
2025-03-18 21:27:34,742 - INFO - Iter 641/10000, Loss: 0.2428, Avg Adaptive LR: 3.296390
2025-03-18 21:27:35,144 - INFO - Iter 651/10000, Loss: 0.2425, Avg Adaptive LR: 3.296370
2025-03-18 21:27:35,554 - INFO - Iter 661/10000, Loss: 0.2422, Avg Adaptive LR: 3.296350
2025-03-18 21:27:35,952 - INFO - Iter 671/10000, Loss: 0.2420, Avg Adaptive LR: 3.296331
2025-03-18 21:27:36,365 - INFO - Iter 681/10000, Loss: 0.2417, Avg Adaptive LR: 3.296312
2025-03-18 21:27:36,767 - INFO - Iter 691/10000, Loss: 0.2414, Avg Adaptive LR: 3.296293
2025-03-18 21:27:37,172 - INFO - Iter 701/10000, Loss: 0.2411, Avg Adaptive LR: 3.296275
2025-03-18 21:27:37,584 - INFO - Iter 711/10000, Loss: 0.2409, Avg Adaptive LR: 3.296257
2025-03-18 21:27:37,979 - INFO - Iter 721/10000, Loss: 0.2406, Avg Adaptive LR: 3.296239
2025-03-18 21:27:38,384 - INFO - Iter 731/10000, Loss: 0.2404, Avg Adaptive LR: 3.296222
2025-03-18 21:27:38,782 - INFO - Iter 741/10000, Loss: 0.2401, Avg Adaptive LR: 3.296205
2025-03-18 21:27:39,176 - INFO - Iter 751/10000, Loss: 0.2399, Avg Adaptive LR: 3.296189
2025-03-18 21:27:39,580 - INFO - Iter 761/10000, Loss: 0.2397, Avg Adaptive LR: 3.296172
2025-03-18 21:27:39,978 - INFO - Iter 771/10000, Loss: 0.2394, Avg Adaptive LR: 3.296156
2025-03-18 21:27:40,387 - INFO - Iter 781/10000, Loss: 0.2392, Avg Adaptive LR: 3.296140
2025-03-18 21:27:40,788 - INFO - Iter 791/10000, Loss: 0.2390, Avg Adaptive LR: 3.296125
2025-03-18 21:27:41,197 - INFO - Iter 801/10000, Loss: 0.2387, Avg Adaptive LR: 3.296109
2025-03-18 21:27:41,603 - INFO - Iter 811/10000, Loss: 0.2385, Avg Adaptive LR: 3.296094
2025-03-18 21:27:42,005 - INFO - Iter 821/10000, Loss: 0.2383, Avg Adaptive LR: 3.296079
2025-03-18 21:27:42,408 - INFO - Iter 831/10000, Loss: 0.2381, Avg Adaptive LR: 3.296065
2025-03-18 21:27:42,808 - INFO - Iter 841/10000, Loss: 0.2379, Avg Adaptive LR: 3.296051
2025-03-18 21:27:43,205 - INFO - Iter 851/10000, Loss: 0.2377, Avg Adaptive LR: 3.296036
2025-03-18 21:27:43,611 - INFO - Iter 861/10000, Loss: 0.2375, Avg Adaptive LR: 3.296022
2025-03-18 21:27:44,017 - INFO - Iter 871/10000, Loss: 0.2373, Avg Adaptive LR: 3.296009
2025-03-18 21:27:44,427 - INFO - Iter 881/10000, Loss: 0.2371, Avg Adaptive LR: 3.295995
2025-03-18 21:27:44,828 - INFO - Iter 891/10000, Loss: 0.2369, Avg Adaptive LR: 3.295982
```

```
2025-03-18 21:27:45,250 - INFO - Iter 901/10000, Loss: 0.2367, Avg Adaptive LR: 3.295969
2025-03-18 21:27:45,666 - INFO - Iter 911/10000, Loss: 0.2365, Avg Adaptive LR: 3.295956
2025-03-18 21:27:46,063 - INFO - Iter 921/10000, Loss: 0.2363, Avg Adaptive LR: 3.295943
2025-03-18 21:27:46,470 - INFO - Iter 931/10000, Loss: 0.2362, Avg Adaptive LR: 3.295930
2025-03-18 21:27:46,865 - INFO - Iter 941/10000, Loss: 0.2360, Avg Adaptive LR: 3.295918
2025-03-18 21:27:47,269 - INFO - Iter 951/10000, Loss: 0.2358, Avg Adaptive LR: 3.295906
2025-03-18 21:27:47,675 - INFO - Iter 961/10000, Loss: 0.2356, Avg Adaptive LR: 3.295894
2025-03-18 21:27:48,080 - INFO - Iter 971/10000, Loss: 0.2355, Avg Adaptive LR: 3.295882
2025-03-18 21:27:48,495 - INFO - Iter 981/10000, Loss: 0.2353, Avg Adaptive LR: 3.295870
2025-03-18 21:27:48,890 - INFO - Iter 991/10000, Loss: 0.2351, Avg Adaptive LR: 3.295858
2025-03-18 21:27:49,296 - INFO - Iter 1001/10000, Loss: 0.2350, Avg Adaptive LR: 3.295847
2025-03-18 21:27:49,699 - INFO - Iter 1011/10000, Loss: 0.2348, Avg Adaptive LR: 3.295836
2025-03-18 21:27:50,103 - INFO - Iter 1021/10000, Loss: 0.2346, Avg Adaptive LR: 3.295825
2025-03-18 21:27:50,516 - INFO - Iter 1031/10000, Loss: 0.2345, Avg Adaptive LR: 3.295814
2025-03-18 21:27:50,916 - INFO - Iter 1041/10000, Loss: 0.2343, Avg Adaptive LR: 3.295803
2025-03-18 21:27:51,311 - INFO - Iter 1051/10000, Loss: 0.2342, Avg Adaptive LR: 3.295792
2025-03-18 21:27:51,704 - INFO - Iter 1061/10000, Loss: 0.2340, Avg Adaptive LR: 3.295781
2025-03-18 21:27:52,103 - INFO - Iter 1071/10000, Loss: 0.2339, Avg Adaptive LR: 3.295771
2025-03-18 21:27:52,516 - INFO - Iter 1081/10000, Loss: 0.2337, Avg Adaptive LR: 3.295760
2025-03-18 21:27:52,918 - INFO - Iter 1091/10000, Loss: 0.2336, Avg Adaptive LR: 3.295750
2025-03-18 21:27:53,320 - INFO - Iter 1101/10000, Loss: 0.2334, Avg Adaptive LR: 3.295740
2025-03-18 21:27:53,715 - INFO - Iter 1111/10000, Loss: 0.2333, Avg Adaptive LR: 3.295730
2025-03-18 21:27:54,117 - INFO - Iter 1121/10000, Loss: 0.2331, Avg Adaptive LR: 3.295720
2025-03-18 21:27:54,522 - INFO - Iter 1131/10000, Loss: 0.2330, Avg Adaptive LR: 3.295710
2025-03-18 21:27:54,920 - INFO - Iter 1141/10000, Loss: 0.2328, Avg Adaptive LR: 3.295701
2025-03-18 21:27:55,350 - INFO - Iter 1151/10000, Loss: 0.2327, Avg Adaptive LR: 3.295691
2025-03-18 21:27:55,762 - INFO - Iter 1161/10000, Loss: 0.2326, Avg Adaptive LR: 3.295682
2025-03-18 21:27:56,174 - INFO - Iter 1171/10000, Loss: 0.2324, Avg Adaptive LR: 3.295672
2025-03-18 21:27:56,585 - INFO - Iter 1181/10000, Loss: 0.2323, Avg Adaptive LR: 3.295663
2025-03-18 21:27:56,983 - INFO - Iter 1191/10000, Loss: 0.2322, Avg Adaptive LR: 3.295654
2025-03-18 21:27:57,402 - INFO - Iter 1201/10000, Loss: 0.2320, Avg Adaptive LR: 3.295645
2025-03-18 21:27:57,804 - INFO - Iter 1211/10000, Loss: 0.2319, Avg Adaptive LR: 3.295636
2025-03-18 21:27:58,214 - INFO - Iter 1221/10000, Loss: 0.2318, Avg Adaptive LR: 3.295627
2025-03-18 21:27:58,631 - INFO - Iter 1231/10000, Loss: 0.2316, Avg Adaptive LR: 3.295618
2025-03-18 21:27:59,038 - INFO - Iter 1241/10000, Loss: 0.2315, Avg Adaptive LR: 3.295610
2025-03-18 21:27:59,447 - INFO - Iter 1251/10000, Loss: 0.2314, Avg Adaptive LR: 3.295601
2025-03-18 21:27:59,847 - INFO - Iter 1261/10000, Loss: 0.2313, Avg Adaptive LR: 3.295593
2025-03-18 21:28:00,255 - INFO - Iter 1271/10000, Loss: 0.2312, Avg Adaptive LR: 3.295584
2025-03-18 21:28:00,675 - INFO - Iter 1281/10000, Loss: 0.2310, Avg Adaptive LR: 3.295576
2025-03-18 21:28:01,081 - INFO - Iter 1291/10000, Loss: 0.2309, Avg Adaptive LR: 3.295568
2025-03-18 21:28:01,489 - INFO - Iter 1301/10000, Loss: 0.2308, Avg Adaptive LR: 3.295560
2025-03-18 21:28:01,892 - INFO - Iter 1311/10000, Loss: 0.2307, Avg Adaptive LR: 3.295551
2025-03-18 21:28:02,300 - INFO - Iter 1321/10000, Loss: 0.2306, Avg Adaptive LR: 3.295543
2025-03-18 21:28:02,704 - INFO - Iter 1331/10000, Loss: 0.2304, Avg Adaptive LR: 3.295536
2025-03-18 21:28:03,108 - INFO - Iter 1341/10000, Loss: 0.2303, Avg Adaptive LR: 3.295528
2025-03-18 21:28:03,539 - INFO - Iter 1351/10000, Loss: 0.2302, Avg Adaptive LR: 3.295520
2025-03-18 21:28:03,962 - INFO - Iter 1361/10000, Loss: 0.2301, Avg Adaptive LR: 3.295512
2025-03-18 21:28:04,373 - INFO - Iter 1371/10000, Loss: 0.2300, Avg Adaptive LR: 3.295505
2025-03-18 21:28:04,771 - INFO - Iter 1381/10000, Loss: 0.2299, Avg Adaptive LR: 3.295497
2025-03-18 21:28:05,173 - INFO - Iter 1391/10000, Loss: 0.2298, Avg Adaptive LR: 3.295490
2025-03-18 21:28:05,578 - INFO - Iter 1401/10000, Loss: 0.2297, Avg Adaptive LR: 3.295482
2025-03-18 21:28:05,983 - INFO - Iter 1411/10000, Loss: 0.2296, Avg Adaptive LR: 3.295475
2025-03-18 21:28:06,392 - INFO - Iter 1421/10000, Loss: 0.2295, Avg Adaptive LR: 3.295468
2025-03-18 21:28:06,796 - INFO - Iter 1431/10000, Loss: 0.2294, Avg Adaptive LR: 3.295460
2025-03-18 21:28:07,189 - INFO - Iter 1441/10000, Loss: 0.2293, Avg Adaptive LR: 3.295453
2025-03-18 21:28:07,597 - INFO - Iter 1451/10000, Loss: 0.2292, Avg Adaptive LR: 3.295446
2025-03-18 21:28:07,990 - INFO - Iter 1461/10000, Loss: 0.2291, Avg Adaptive LR: 3.295439
2025-03-18 21:28:08,399 - INFO - Iter 1471/10000, Loss: 0.2290, Avg Adaptive LR: 3.295432
2025-03-18 21:28:08,792 - INFO - Iter 1481/10000, Loss: 0.2289, Avg Adaptive LR: 3.295425
2025-03-18 21:28:09,188 - INFO - Iter 1491/10000, Loss: 0.2288, Avg Adaptive LR: 3.295418
2025-03-18 21:28:09,586 - INFO - Iter 1501/10000, Loss: 0.2287, Avg Adaptive LR: 3.295412
2025-03-18 21:28:09,985 - INFO - Iter 1511/10000, Loss: 0.2286, Avg Adaptive LR: 3.295405
2025-03-18 21:28:10,382 - INFO - Iter 1521/10000, Loss: 0.2285, Avg Adaptive LR: 3.295398
2025-03-18 21:28:10,774 - INFO - Iter 1531/10000, Loss: 0.2284, Avg Adaptive LR: 3.295392
2025-03-18 21:28:11,170 - INFO - Iter 1541/10000, Loss: 0.2283, Avg Adaptive LR: 3.295385
2025-03-18 21:28:11,573 - INFO - Iter 1551/10000, Loss: 0.2282, Avg Adaptive LR: 3.295379
2025-03-18 21:28:11,966 - INFO - Iter 1561/10000, Loss: 0.2281, Avg Adaptive LR: 3.295372
2025-03-18 21:28:12,365 - INFO - Iter 1571/10000, Loss: 0.2280, Avg Adaptive LR: 3.295366
2025-03-18 21:28:12,757 - INFO - Iter 1581/10000, Loss: 0.2279, Avg Adaptive LR: 3.295360
2025-03-18 21:28:13,151 - INFO - Iter 1591/10000, Loss: 0.2278, Avg Adaptive LR: 3.295353
2025-03-18 21:28:13,552 - INFO - Iter 1601/10000, Loss: 0.2277, Avg Adaptive LR: 3.295347
2025-03-18 21:28:13,945 - INFO - Iter 1611/10000, Loss: 0.2276, Avg Adaptive LR: 3.295341
```

```
2025-03-18 21:28:14,366 - INFO - Iter 1621/10000, Loss: 0.2275, Avg Adaptive LR: 3.295335
2025-03-18 21:28:14,760 - INFO - Iter 1631/10000, Loss: 0.2275, Avg Adaptive LR: 3.295329
2025-03-18 21:28:15,150 - INFO - Iter 1641/10000, Loss: 0.2274, Avg Adaptive LR: 3.295323
2025-03-18 21:28:15,547 - INFO - Iter 1651/10000, Loss: 0.2273, Avg Adaptive LR: 3.295317
2025-03-18 21:28:15,943 - INFO - Iter 1661/10000, Loss: 0.2272, Avg Adaptive LR: 3.295311
2025-03-18 21:28:16,348 - INFO - Iter 1671/10000, Loss: 0.2271, Avg Adaptive LR: 3.295305
2025-03-18 21:28:16,739 - INFO - Iter 1681/10000, Loss: 0.2270, Avg Adaptive LR: 3.295299
2025-03-18 21:28:17,139 - INFO - Iter 1691/10000, Loss: 0.2269, Avg Adaptive LR: 3.295293
2025-03-18 21:28:17,535 - INFO - Iter 1701/10000, Loss: 0.2269, Avg Adaptive LR: 3.295287
2025-03-18 21:28:17,932 - INFO - Iter 1711/10000, Loss: 0.2268, Avg Adaptive LR: 3.295282
2025-03-18 21:28:18,331 - INFO - Iter 1721/10000, Loss: 0.2267, Avg Adaptive LR: 3.295276
2025-03-18 21:28:18,725 - INFO - Iter 1731/10000, Loss: 0.2266, Avg Adaptive LR: 3.295270
2025-03-18 21:28:19,123 - INFO - Iter 1741/10000, Loss: 0.2265, Avg Adaptive LR: 3.295265
2025-03-18 21:28:19,524 - INFO - Iter 1751/10000, Loss: 0.2265, Avg Adaptive LR: 3.295259
2025-03-18 21:28:19,921 - INFO - Iter 1761/10000, Loss: 0.2264, Avg Adaptive LR: 3.295254
2025-03-18 21:28:20,326 - INFO - Iter 1771/10000, Loss: 0.2263, Avg Adaptive LR: 3.295248
2025-03-18 21:28:20,723 - INFO - Iter 1781/10000, Loss: 0.2262, Avg Adaptive LR: 3.295243
2025-03-18 21:28:21,121 - INFO - Iter 1791/10000, Loss: 0.2261, Avg Adaptive LR: 3.295238
2025-03-18 21:28:21,522 - INFO - Iter 1801/10000, Loss: 0.2261, Avg Adaptive LR: 3.295232
2025-03-18 21:28:21,917 - INFO - Iter 1811/10000, Loss: 0.2260, Avg Adaptive LR: 3.295227
2025-03-18 21:28:22,317 - INFO - Iter 1821/10000, Loss: 0.2259, Avg Adaptive LR: 3.295222
2025-03-18 21:28:22,713 - INFO - Iter 1831/10000, Loss: 0.2258, Avg Adaptive LR: 3.295216
2025-03-18 21:28:23,107 - INFO - Iter 1841/10000, Loss: 0.2258, Avg Adaptive LR: 3.295211
2025-03-18 21:28:23,515 - INFO - Iter 1851/10000, Loss: 0.2257, Avg Adaptive LR: 3.295206
2025-03-18 21:28:23,905 - INFO - Iter 1861/10000, Loss: 0.2256, Avg Adaptive LR: 3.295201
2025-03-18 21:28:24,306 - INFO - Iter 1871/10000, Loss: 0.2255, Avg Adaptive LR: 3.295196
2025-03-18 21:28:24,720 - INFO - Iter 1881/10000, Loss: 0.2255, Avg Adaptive LR: 3.295191
2025-03-18 21:28:25,118 - INFO - Iter 1891/10000, Loss: 0.2254, Avg Adaptive LR: 3.295186
2025-03-18 21:28:25,520 - INFO - Iter 1901/10000, Loss: 0.2253, Avg Adaptive LR: 3.295181
2025-03-18 21:28:25,919 - INFO - Iter 1911/10000, Loss: 0.2252, Avg Adaptive LR: 3.295176
2025-03-18 21:28:26,341 - INFO - Iter 1921/10000, Loss: 0.2252, Avg Adaptive LR: 3.295171
2025-03-18 21:28:26,748 - INFO - Iter 1931/10000, Loss: 0.2251, Avg Adaptive LR: 3.295166
2025-03-18 21:28:27,142 - INFO - Iter 1941/10000, Loss: 0.2250, Avg Adaptive LR: 3.295161
2025-03-18 21:28:27,548 - INFO - Iter 1951/10000, Loss: 0.2250, Avg Adaptive LR: 3.295157
2025-03-18 21:28:27,943 - INFO - Iter 1961/10000, Loss: 0.2249, Avg Adaptive LR: 3.295152
2025-03-18 21:28:28,347 - INFO - Iter 1971/10000, Loss: 0.2248, Avg Adaptive LR: 3.295147
2025-03-18 21:28:28,737 - INFO - Iter 1981/10000, Loss: 0.2248, Avg Adaptive LR: 3.295142
2025-03-18 21:28:29,129 - INFO - Iter 1991/10000, Loss: 0.2247, Avg Adaptive LR: 3.295138
2025-03-18 21:28:29,531 - INFO - Iter 2001/10000, Loss: 0.2246, Avg Adaptive LR: 3.295133
2025-03-18 21:28:29,923 - INFO - Iter 2011/10000, Loss: 0.2246, Avg Adaptive LR: 3.295128
2025-03-18 21:28:30,323 - INFO - Iter 2021/10000, Loss: 0.2245, Avg Adaptive LR: 3.295124
2025-03-18 21:28:30,720 - INFO - Iter 2031/10000, Loss: 0.2244, Avg Adaptive LR: 3.295119
2025-03-18 21:28:31,114 - INFO - Iter 2041/10000, Loss: 0.2244, Avg Adaptive LR: 3.295115
2025-03-18 21:28:31,518 - INFO - Iter 2051/10000, Loss: 0.2243, Avg Adaptive LR: 3.295110
2025-03-18 21:28:31,908 - INFO - Iter 2061/10000, Loss: 0.2242, Avg Adaptive LR: 3.295106
2025-03-18 21:28:32,308 - INFO - Iter 2071/10000, Loss: 0.2242, Avg Adaptive LR: 3.295101
2025-03-18 21:28:32,705 - INFO - Iter 2081/10000, Loss: 0.2241, Avg Adaptive LR: 3.295097
2025-03-18 21:28:33,102 - INFO - Iter 2091/10000, Loss: 0.2240, Avg Adaptive LR: 3.295093
2025-03-18 21:28:33,500 - INFO - Iter 2101/10000, Loss: 0.2240, Avg Adaptive LR: 3.295088
2025-03-18 21:28:33,891 - INFO - Iter 2111/10000, Loss: 0.2239, Avg Adaptive LR: 3.295084
2025-03-18 21:28:34,296 - INFO - Iter 2121/10000, Loss: 0.2239, Avg Adaptive LR: 3.295080
2025-03-18 21:28:34,686 - INFO - Iter 2131/10000, Loss: 0.2238, Avg Adaptive LR: 3.295075
2025-03-18 21:28:35,101 - INFO - Iter 2141/10000, Loss: 0.2237, Avg Adaptive LR: 3.295071
2025-03-18 21:28:35,496 - INFO - Iter 2151/10000, Loss: 0.2237, Avg Adaptive LR: 3.295067
2025-03-18 21:28:35,893 - INFO - Iter 2161/10000, Loss: 0.2236, Avg Adaptive LR: 3.295063
2025-03-18 21:28:36,297 - INFO - Iter 2171/10000, Loss: 0.2235, Avg Adaptive LR: 3.295058
2025-03-18 21:28:36,696 - INFO - Iter 2181/10000, Loss: 0.2235, Avg Adaptive LR: 3.295054
2025-03-18 21:28:37,090 - INFO - Iter 2191/10000, Loss: 0.2234, Avg Adaptive LR: 3.295050
2025-03-18 21:28:37,494 - INFO - Iter 2201/10000, Loss: 0.2234, Avg Adaptive LR: 3.295046
2025-03-18 21:28:37,889 - INFO - Iter 2211/10000, Loss: 0.2233, Avg Adaptive LR: 3.295042
2025-03-18 21:28:38,287 - INFO - Iter 2221/10000, Loss: 0.2233, Avg Adaptive LR: 3.295038
2025-03-18 21:28:38,682 - INFO - Iter 2231/10000, Loss: 0.2232, Avg Adaptive LR: 3.295034
2025-03-18 21:28:39,071 - INFO - Iter 2241/10000, Loss: 0.2231, Avg Adaptive LR: 3.295030
2025-03-18 21:28:39,469 - INFO - Iter 2251/10000, Loss: 0.2231, Avg Adaptive LR: 3.295026
2025-03-18 21:28:39,866 - INFO - Iter 2261/10000, Loss: 0.2230, Avg Adaptive LR: 3.295022
2025-03-18 21:28:40,263 - INFO - Iter 2271/10000, Loss: 0.2230, Avg Adaptive LR: 3.295018
2025-03-18 21:28:40,654 - INFO - Iter 2281/10000, Loss: 0.2229, Avg Adaptive LR: 3.295014
2025-03-18 21:28:41,045 - INFO - Iter 2291/10000, Loss: 0.2229, Avg Adaptive LR: 3.295010
2025-03-18 21:28:41,448 - INFO - Iter 2301/10000, Loss: 0.2228, Avg Adaptive LR: 3.295007
2025-03-18 21:28:41,847 - INFO - Iter 2311/10000, Loss: 0.2227, Avg Adaptive LR: 3.295003
2025-03-18 21:28:42,250 - INFO - Iter 2321/10000, Loss: 0.2227, Avg Adaptive LR: 3.294999
2025-03-18 21:28:42,645 - INFO - Iter 2331/10000, Loss: 0.2226, Avg Adaptive LR: 3.294995
```

```
2025-03-18 21:28:43,042 - INFO - Iter 2341/10000, Loss: 0.2226, Avg Adaptive LR: 3.294991
2025-03-18 21:28:43,446 - INFO - Iter 2351/10000, Loss: 0.2225, Avg Adaptive LR: 3.294988
2025-03-18 21:28:43,842 - INFO - Iter 2361/10000, Loss: 0.2225, Avg Adaptive LR: 3.294984
2025-03-18 21:28:44,243 - INFO - Iter 2371/10000, Loss: 0.2224, Avg Adaptive LR: 3.294980
2025-03-18 21:28:44,643 - INFO - Iter 2381/10000, Loss: 0.2224, Avg Adaptive LR: 3.294976
2025-03-18 21:28:45,038 - INFO - Iter 2391/10000, Loss: 0.2223, Avg Adaptive LR: 3.294973
2025-03-18 21:28:45,460 - INFO - Iter 2401/10000, Loss: 0.2223, Avg Adaptive LR: 3.294969
2025-03-18 21:28:45,852 - INFO - Iter 2411/10000, Loss: 0.2222, Avg Adaptive LR: 3.294966
2025-03-18 21:28:46,250 - INFO - Iter 2421/10000, Loss: 0.2222, Avg Adaptive LR: 3.294962
2025-03-18 21:28:46,662 - INFO - Iter 2431/10000, Loss: 0.2221, Avg Adaptive LR: 3.294958
2025-03-18 21:28:47,051 - INFO - Iter 2441/10000, Loss: 0.2220, Avg Adaptive LR: 3.294955
2025-03-18 21:28:47,451 - INFO - Iter 2451/10000, Loss: 0.2220, Avg Adaptive LR: 3.294951
2025-03-18 21:28:47,847 - INFO - Iter 2461/10000, Loss: 0.2219, Avg Adaptive LR: 3.294948
2025-03-18 21:28:48,244 - INFO - Iter 2471/10000, Loss: 0.2219, Avg Adaptive LR: 3.294944
2025-03-18 21:28:48,639 - INFO - Iter 2481/10000, Loss: 0.2218, Avg Adaptive LR: 3.294941
2025-03-18 21:28:49,030 - INFO - Iter 2491/10000, Loss: 0.2218, Avg Adaptive LR: 3.294937
2025-03-18 21:28:49,432 - INFO - Iter 2501/10000, Loss: 0.2217, Avg Adaptive LR: 3.294934
2025-03-18 21:28:49,827 - INFO - Iter 2511/10000, Loss: 0.2217, Avg Adaptive LR: 3.294930
2025-03-18 21:28:50,221 - INFO - Iter 2521/10000, Loss: 0.2216, Avg Adaptive LR: 3.294927
2025-03-18 21:28:50,625 - INFO - Iter 2531/10000, Loss: 0.2216, Avg Adaptive LR: 3.294924
2025-03-18 21:28:51,018 - INFO - Iter 2541/10000, Loss: 0.2215, Avg Adaptive LR: 3.294920
2025-03-18 21:28:51,425 - INFO - Iter 2551/10000, Loss: 0.2215, Avg Adaptive LR: 3.294917
2025-03-18 21:28:51,823 - INFO - Iter 2561/10000, Loss: 0.2215, Avg Adaptive LR: 3.294914
2025-03-18 21:28:52,218 - INFO - Iter 2571/10000, Loss: 0.2214, Avg Adaptive LR: 3.294910
2025-03-18 21:28:52,618 - INFO - Iter 2581/10000, Loss: 0.2214, Avg Adaptive LR: 3.294907
2025-03-18 21:28:53,010 - INFO - Iter 2591/10000, Loss: 0.2213, Avg Adaptive LR: 3.294904
2025-03-18 21:28:53,410 - INFO - Iter 2601/10000, Loss: 0.2213, Avg Adaptive LR: 3.294900
2025-03-18 21:28:53,807 - INFO - Iter 2611/10000, Loss: 0.2212, Avg Adaptive LR: 3.294897
2025-03-18 21:28:54,202 - INFO - Iter 2621/10000, Loss: 0.2212, Avg Adaptive LR: 3.294894
2025-03-18 21:28:54,601 - INFO - Iter 2631/10000, Loss: 0.2211, Avg Adaptive LR: 3.294891
2025-03-18 21:28:55,001 - INFO - Iter 2641/10000, Loss: 0.2211, Avg Adaptive LR: 3.294887
2025-03-18 21:28:55,402 - INFO - Iter 2651/10000, Loss: 0.2210, Avg Adaptive LR: 3.294884
2025-03-18 21:28:55,814 - INFO - Iter 2661/10000, Loss: 0.2210, Avg Adaptive LR: 3.294881
2025-03-18 21:28:56,207 - INFO - Iter 2671/10000, Loss: 0.2209, Avg Adaptive LR: 3.294878
2025-03-18 21:28:56,612 - INFO - Iter 2681/10000, Loss: 0.2209, Avg Adaptive LR: 3.294875
2025-03-18 21:28:57,017 - INFO - Iter 2691/10000, Loss: 0.2208, Avg Adaptive LR: 3.294872
2025-03-18 21:28:57,416 - INFO - Iter 2701/10000, Loss: 0.2208, Avg Adaptive LR: 3.294869
2025-03-18 21:28:57,815 - INFO - Iter 2711/10000, Loss: 0.2208, Avg Adaptive LR: 3.294866
2025-03-18 21:28:58,212 - INFO - Iter 2721/10000, Loss: 0.2207, Avg Adaptive LR: 3.294862
2025-03-18 21:28:58,618 - INFO - Iter 2731/10000, Loss: 0.2207, Avg Adaptive LR: 3.294859
2025-03-18 21:28:59,008 - INFO - Iter 2741/10000, Loss: 0.2206, Avg Adaptive LR: 3.294856
2025-03-18 21:28:59,405 - INFO - Iter 2751/10000, Loss: 0.2206, Avg Adaptive LR: 3.294853
2025-03-18 21:28:59,800 - INFO - Iter 2761/10000, Loss: 0.2205, Avg Adaptive LR: 3.294850
2025-03-18 21:29:00,196 - INFO - Iter 2771/10000, Loss: 0.2205, Avg Adaptive LR: 3.294847
2025-03-18 21:29:00,601 - INFO - Iter 2781/10000, Loss: 0.2204, Avg Adaptive LR: 3.294844
2025-03-18 21:29:00,995 - INFO - Iter 2791/10000, Loss: 0.2204, Avg Adaptive LR: 3.294841
2025-03-18 21:29:01,398 - INFO - Iter 2801/10000, Loss: 0.2204, Avg Adaptive LR: 3.294838
2025-03-18 21:29:01,796 - INFO - Iter 2811/10000, Loss: 0.2203, Avg Adaptive LR: 3.294835
2025-03-18 21:29:02,198 - INFO - Iter 2821/10000, Loss: 0.2203, Avg Adaptive LR: 3.294833
2025-03-18 21:29:02,602 - INFO - Iter 2831/10000, Loss: 0.2202, Avg Adaptive LR: 3.294830
2025-03-18 21:29:02,998 - INFO - Iter 2841/10000, Loss: 0.2202, Avg Adaptive LR: 3.294827
2025-03-18 21:29:03,395 - INFO - Iter 2851/10000, Loss: 0.2202, Avg Adaptive LR: 3.294824
2025-03-18 21:29:03,793 - INFO - Iter 2861/10000, Loss: 0.2201, Avg Adaptive LR: 3.294821
2025-03-18 21:29:04,184 - INFO - Iter 2871/10000, Loss: 0.2201, Avg Adaptive LR: 3.294818
2025-03-18 21:29:04,583 - INFO - Iter 2881/10000, Loss: 0.2200, Avg Adaptive LR: 3.294815
2025-03-18 21:29:04,980 - INFO - Iter 2891/10000, Loss: 0.2200, Avg Adaptive LR: 3.294812
2025-03-18 21:29:05,382 - INFO - Iter 2901/10000, Loss: 0.2199, Avg Adaptive LR: 3.294810
2025-03-18 21:29:05,799 - INFO - Iter 2911/10000, Loss: 0.2199, Avg Adaptive LR: 3.294807
2025-03-18 21:29:06,202 - INFO - Iter 2921/10000, Loss: 0.2199, Avg Adaptive LR: 3.294804
2025-03-18 21:29:06,620 - INFO - Iter 2931/10000, Loss: 0.2198, Avg Adaptive LR: 3.294801
2025-03-18 21:29:07,035 - INFO - Iter 2941/10000, Loss: 0.2198, Avg Adaptive LR: 3.294799
2025-03-18 21:29:07,440 - INFO - Iter 2951/10000, Loss: 0.2197, Avg Adaptive LR: 3.294796
2025-03-18 21:29:07,833 - INFO - Iter 2961/10000, Loss: 0.2197, Avg Adaptive LR: 3.294793
2025-03-18 21:29:08,227 - INFO - Iter 2971/10000, Loss: 0.2197, Avg Adaptive LR: 3.294790
2025-03-18 21:29:08,632 - INFO - Iter 2981/10000, Loss: 0.2196, Avg Adaptive LR: 3.294788
2025-03-18 21:29:09,036 - INFO - Iter 2991/10000, Loss: 0.2196, Avg Adaptive LR: 3.294785
2025-03-18 21:29:09,438 - INFO - Iter 3001/10000, Loss: 0.2196, Avg Adaptive LR: 3.294782
2025-03-18 21:29:09,832 - INFO - Iter 3011/10000, Loss: 0.2195, Avg Adaptive LR: 3.294780
2025-03-18 21:29:10,223 - INFO - Iter 3021/10000, Loss: 0.2195, Avg Adaptive LR: 3.294777
2025-03-18 21:29:10,628 - INFO - Iter 3031/10000, Loss: 0.2194, Avg Adaptive LR: 3.294774
2025-03-18 21:29:11,019 - INFO - Iter 3041/10000, Loss: 0.2194, Avg Adaptive LR: 3.294772
2025-03-18 21:29:11,422 - INFO - Iter 3051/10000, Loss: 0.2194, Avg Adaptive LR: 3.294769
```

```
2025-03-18 21:29:11,820 - INFO - Iter 3061/10000, Loss: 0.2193, Avg Adaptive LR: 3.294766
2025-03-18 21:29:12,213 - INFO - Iter 3071/10000, Loss: 0.2193, Avg Adaptive LR: 3.294764
2025-03-18 21:29:12,624 - INFO - Iter 3081/10000, Loss: 0.2192, Avg Adaptive LR: 3.294761
2025-03-18 21:29:13,015 - INFO - Iter 3091/10000, Loss: 0.2192, Avg Adaptive LR: 3.294759
2025-03-18 21:29:13,415 - INFO - Iter 3101/10000, Loss: 0.2192, Avg Adaptive LR: 3.294756
2025-03-18 21:29:13,811 - INFO - Iter 3111/10000, Loss: 0.2191, Avg Adaptive LR: 3.294753
2025-03-18 21:29:14,209 - INFO - Iter 3121/10000, Loss: 0.2191, Avg Adaptive LR: 3.294751
2025-03-18 21:29:14,613 - INFO - Iter 3131/10000, Loss: 0.2191, Avg Adaptive LR: 3.294748
2025-03-18 21:29:15,008 - INFO - Iter 3141/10000, Loss: 0.2190, Avg Adaptive LR: 3.294746
2025-03-18 21:29:15,408 - INFO - Iter 3151/10000, Loss: 0.2190, Avg Adaptive LR: 3.294743
2025-03-18 21:29:15,802 - INFO - Iter 3161/10000, Loss: 0.2190, Avg Adaptive LR: 3.294741
2025-03-18 21:29:16,216 - INFO - Iter 3171/10000, Loss: 0.2189, Avg Adaptive LR: 3.294738
2025-03-18 21:29:16,617 - INFO - Iter 3181/10000, Loss: 0.2189, Avg Adaptive LR: 3.294736
2025-03-18 21:29:17,021 - INFO - Iter 3191/10000, Loss: 0.2188, Avg Adaptive LR: 3.294733
2025-03-18 21:29:17,422 - INFO - Iter 3201/10000, Loss: 0.2188, Avg Adaptive LR: 3.294731
2025-03-18 21:29:17,812 - INFO - Iter 3211/10000, Loss: 0.2188, Avg Adaptive LR: 3.294729
2025-03-18 21:29:18,210 - INFO - Iter 3221/10000, Loss: 0.2187, Avg Adaptive LR: 3.294726
2025-03-18 21:29:18,612 - INFO - Iter 3231/10000, Loss: 0.2187, Avg Adaptive LR: 3.294724
2025-03-18 21:29:19,002 - INFO - Iter 3241/10000, Loss: 0.2187, Avg Adaptive LR: 3.294721
2025-03-18 21:29:19,405 - INFO - Iter 3251/10000, Loss: 0.2186, Avg Adaptive LR: 3.294719
2025-03-18 21:29:19,802 - INFO - Iter 3261/10000, Loss: 0.2186, Avg Adaptive LR: 3.294716
2025-03-18 21:29:20,193 - INFO - Iter 3271/10000, Loss: 0.2186, Avg Adaptive LR: 3.294714
2025-03-18 21:29:20,596 - INFO - Iter 3281/10000, Loss: 0.2185, Avg Adaptive LR: 3.294712
2025-03-18 21:29:20,990 - INFO - Iter 3291/10000, Loss: 0.2185, Avg Adaptive LR: 3.294709
2025-03-18 21:29:21,396 - INFO - Iter 3301/10000, Loss: 0.2185, Avg Adaptive LR: 3.294707
2025-03-18 21:29:21,788 - INFO - Iter 3311/10000, Loss: 0.2184, Avg Adaptive LR: 3.294705
2025-03-18 21:29:22,180 - INFO - Iter 3321/10000, Loss: 0.2184, Avg Adaptive LR: 3.294702
2025-03-18 21:29:22,579 - INFO - Iter 3331/10000, Loss: 0.2184, Avg Adaptive LR: 3.294700
2025-03-18 21:29:22,973 - INFO - Iter 3341/10000, Loss: 0.2183, Avg Adaptive LR: 3.294698
2025-03-18 21:29:23,369 - INFO - Iter 3351/10000, Loss: 0.2183, Avg Adaptive LR: 3.294695
2025-03-18 21:29:23,763 - INFO - Iter 3361/10000, Loss: 0.2183, Avg Adaptive LR: 3.294693
2025-03-18 21:29:24,155 - INFO - Iter 3371/10000, Loss: 0.2182, Avg Adaptive LR: 3.294691
2025-03-18 21:29:24,557 - INFO - Iter 3381/10000, Loss: 0.2182, Avg Adaptive LR: 3.294689
2025-03-18 21:29:24,949 - INFO - Iter 3391/10000, Loss: 0.2182, Avg Adaptive LR: 3.294686
2025-03-18 21:29:25,351 - INFO - Iter 3401/10000, Loss: 0.2181, Avg Adaptive LR: 3.294684
2025-03-18 21:29:25,744 - INFO - Iter 3411/10000, Loss: 0.2181, Avg Adaptive LR: 3.294682
2025-03-18 21:29:26,137 - INFO - Iter 3421/10000, Loss: 0.2181, Avg Adaptive LR: 3.294680
2025-03-18 21:29:26,556 - INFO - Iter 3431/10000, Loss: 0.2180, Avg Adaptive LR: 3.294677
2025-03-18 21:29:26,945 - INFO - Iter 3441/10000, Loss: 0.2180, Avg Adaptive LR: 3.294675
2025-03-18 21:29:27,356 - INFO - Iter 3451/10000, Loss: 0.2180, Avg Adaptive LR: 3.294673
2025-03-18 21:29:27,761 - INFO - Iter 3461/10000, Loss: 0.2179, Avg Adaptive LR: 3.294671
2025-03-18 21:29:28,159 - INFO - Iter 3471/10000, Loss: 0.2179, Avg Adaptive LR: 3.294668
2025-03-18 21:29:28,563 - INFO - Iter 3481/10000, Loss: 0.2179, Avg Adaptive LR: 3.294666
2025-03-18 21:29:28,955 - INFO - Iter 3491/10000, Loss: 0.2178, Avg Adaptive LR: 3.294664
2025-03-18 21:29:29,362 - INFO - Iter 3501/10000, Loss: 0.2178, Avg Adaptive LR: 3.294662
2025-03-18 21:29:29,750 - INFO - Iter 3511/10000, Loss: 0.2178, Avg Adaptive LR: 3.294660
2025-03-18 21:29:30,146 - INFO - Iter 3521/10000, Loss: 0.2177, Avg Adaptive LR: 3.294658
2025-03-18 21:29:30,550 - INFO - Iter 3531/10000, Loss: 0.2177, Avg Adaptive LR: 3.294655
2025-03-18 21:29:30,944 - INFO - Iter 3541/10000, Loss: 0.2177, Avg Adaptive LR: 3.294653
2025-03-18 21:29:31,351 - INFO - Iter 3551/10000, Loss: 0.2177, Avg Adaptive LR: 3.294651
2025-03-18 21:29:31,744 - INFO - Iter 3561/10000, Loss: 0.2176, Avg Adaptive LR: 3.294649
2025-03-18 21:29:32,137 - INFO - Iter 3571/10000, Loss: 0.2176, Avg Adaptive LR: 3.294647
2025-03-18 21:29:32,543 - INFO - Iter 3581/10000, Loss: 0.2176, Avg Adaptive LR: 3.294645
2025-03-18 21:29:32,938 - INFO - Iter 3591/10000, Loss: 0.2175, Avg Adaptive LR: 3.294643
2025-03-18 21:29:33,336 - INFO - Iter 3601/10000, Loss: 0.2175, Avg Adaptive LR: 3.294641
2025-03-18 21:29:33,730 - INFO - Iter 3611/10000, Loss: 0.2175, Avg Adaptive LR: 3.294639
2025-03-18 21:29:34,126 - INFO - Iter 3621/10000, Loss: 0.2174, Avg Adaptive LR: 3.294637
2025-03-18 21:29:34,526 - INFO - Iter 3631/10000, Loss: 0.2174, Avg Adaptive LR: 3.294635
2025-03-18 21:29:34,917 - INFO - Iter 3641/10000, Loss: 0.2174, Avg Adaptive LR: 3.294632
2025-03-18 21:29:35,318 - INFO - Iter 3651/10000, Loss: 0.2174, Avg Adaptive LR: 3.294630
2025-03-18 21:29:35,713 - INFO - Iter 3661/10000, Loss: 0.2173, Avg Adaptive LR: 3.294628
2025-03-18 21:29:36,106 - INFO - Iter 3671/10000, Loss: 0.2173, Avg Adaptive LR: 3.294626
2025-03-18 21:29:36,512 - INFO - Iter 3681/10000, Loss: 0.2173, Avg Adaptive LR: 3.294624
2025-03-18 21:29:36,918 - INFO - Iter 3691/10000, Loss: 0.2172, Avg Adaptive LR: 3.294622
2025-03-18 21:29:37,333 - INFO - Iter 3701/10000, Loss: 0.2172, Avg Adaptive LR: 3.294620
2025-03-18 21:29:37,728 - INFO - Iter 3711/10000, Loss: 0.2172, Avg Adaptive LR: 3.294618
2025-03-18 21:29:38,125 - INFO - Iter 3721/10000, Loss: 0.2172, Avg Adaptive LR: 3.294616
2025-03-18 21:29:38,526 - INFO - Iter 3731/10000, Loss: 0.2171, Avg Adaptive LR: 3.294614
2025-03-18 21:29:38,922 - INFO - Iter 3741/10000, Loss: 0.2171, Avg Adaptive LR: 3.294612
2025-03-18 21:29:39,332 - INFO - Iter 3751/10000, Loss: 0.2171, Avg Adaptive LR: 3.294610
2025-03-18 21:29:39,729 - INFO - Iter 3761/10000, Loss: 0.2170, Avg Adaptive LR: 3.294608
2025-03-18 21:29:40,124 - INFO - Iter 3771/10000, Loss: 0.2170, Avg Adaptive LR: 3.294607
```

```
2025-03-18 21:29:40,526 - INFO - Iter 3781/10000, Loss: 0.2170, Avg Adaptive LR: 3.294605
2025-03-18 21:29:40,918 - INFO - Iter 3791/10000, Loss: 0.2170, Avg Adaptive LR: 3.294603
2025-03-18 21:29:41,324 - INFO - Iter 3801/10000, Loss: 0.2169, Avg Adaptive LR: 3.294601
2025-03-18 21:29:41,716 - INFO - Iter 3811/10000, Loss: 0.2169, Avg Adaptive LR: 3.294599
2025-03-18 21:29:42,116 - INFO - Iter 3821/10000, Loss: 0.2169, Avg Adaptive LR: 3.294597
2025-03-18 21:29:42,514 - INFO - Iter 3831/10000, Loss: 0.2168, Avg Adaptive LR: 3.294595
2025-03-18 21:29:42,918 - INFO - Iter 3841/10000, Loss: 0.2168, Avg Adaptive LR: 3.294593
2025-03-18 21:29:43,317 - INFO - Iter 3851/10000, Loss: 0.2168, Avg Adaptive LR: 3.294591
2025-03-18 21:29:43,709 - INFO - Iter 3861/10000, Loss: 0.2168, Avg Adaptive LR: 3.294589
2025-03-18 21:29:44,103 - INFO - Iter 3871/10000, Loss: 0.2167, Avg Adaptive LR: 3.294587
2025-03-18 21:29:44,509 - INFO - Iter 3881/10000, Loss: 0.2167, Avg Adaptive LR: 3.294586
2025-03-18 21:29:44,903 - INFO - Iter 3891/10000, Loss: 0.2167, Avg Adaptive LR: 3.294584
2025-03-18 21:29:45,303 - INFO - Iter 3901/10000, Loss: 0.2167, Avg Adaptive LR: 3.294582
2025-03-18 21:29:45,693 - INFO - Iter 3911/10000, Loss: 0.2166, Avg Adaptive LR: 3.294580
2025-03-18 21:29:46,086 - INFO - Iter 3921/10000, Loss: 0.2166, Avg Adaptive LR: 3.294578
2025-03-18 21:29:46,494 - INFO - Iter 3931/10000, Loss: 0.2166, Avg Adaptive LR: 3.294576
2025-03-18 21:29:46,883 - INFO - Iter 3941/10000, Loss: 0.2165, Avg Adaptive LR: 3.294574
2025-03-18 21:29:47,302 - INFO - Iter 3951/10000, Loss: 0.2165, Avg Adaptive LR: 3.294573
2025-03-18 21:29:47,704 - INFO - Iter 3961/10000, Loss: 0.2165, Avg Adaptive LR: 3.294571
2025-03-18 21:29:48,100 - INFO - Iter 3971/10000, Loss: 0.2165, Avg Adaptive LR: 3.294569
2025-03-18 21:29:48,507 - INFO - Iter 3981/10000, Loss: 0.2164, Avg Adaptive LR: 3.294567
2025-03-18 21:29:48,907 - INFO - Iter 3991/10000, Loss: 0.2164, Avg Adaptive LR: 3.294565
2025-03-18 21:29:49,308 - INFO - Iter 4001/10000, Loss: 0.2164, Avg Adaptive LR: 3.294564
2025-03-18 21:29:49,702 - INFO - Iter 4011/10000, Loss: 0.2164, Avg Adaptive LR: 3.294562
2025-03-18 21:29:50,095 - INFO - Iter 4021/10000, Loss: 0.2163, Avg Adaptive LR: 3.294560
2025-03-18 21:29:50,498 - INFO - Iter 4031/10000, Loss: 0.2163, Avg Adaptive LR: 3.294558
2025-03-18 21:29:50,892 - INFO - Iter 4041/10000, Loss: 0.2163, Avg Adaptive LR: 3.294556
2025-03-18 21:29:51,296 - INFO - Iter 4051/10000, Loss: 0.2163, Avg Adaptive LR: 3.294555
2025-03-18 21:29:51,688 - INFO - Iter 4061/10000, Loss: 0.2162, Avg Adaptive LR: 3.294553
2025-03-18 21:29:52,084 - INFO - Iter 4071/10000, Loss: 0.2162, Avg Adaptive LR: 3.294551
2025-03-18 21:29:52,484 - INFO - Iter 4081/10000, Loss: 0.2162, Avg Adaptive LR: 3.294549
2025-03-18 21:29:52,880 - INFO - Iter 4091/10000, Loss: 0.2162, Avg Adaptive LR: 3.294548
2025-03-18 21:29:53,278 - INFO - Iter 4101/10000, Loss: 0.2161, Avg Adaptive LR: 3.294546
2025-03-18 21:29:53,671 - INFO - Iter 4111/10000, Loss: 0.2161, Avg Adaptive LR: 3.294544
2025-03-18 21:29:54,065 - INFO - Iter 4121/10000, Loss: 0.2161, Avg Adaptive LR: 3.294542
2025-03-18 21:29:54,466 - INFO - Iter 4131/10000, Loss: 0.2161, Avg Adaptive LR: 3.294541
2025-03-18 21:29:54,861 - INFO - Iter 4141/10000, Loss: 0.2160, Avg Adaptive LR: 3.294539
2025-03-18 21:29:55,254 - INFO - Iter 4151/10000, Loss: 0.2160, Avg Adaptive LR: 3.294537
2025-03-18 21:29:55,650 - INFO - Iter 4161/10000, Loss: 0.2160, Avg Adaptive LR: 3.294536
2025-03-18 21:29:56,045 - INFO - Iter 4171/10000, Loss: 0.2160, Avg Adaptive LR: 3.294534
2025-03-18 21:29:56,446 - INFO - Iter 4181/10000, Loss: 0.2159, Avg Adaptive LR: 3.294532
2025-03-18 21:29:56,842 - INFO - Iter 4191/10000, Loss: 0.2159, Avg Adaptive LR: 3.294531
2025-03-18 21:29:57,236 - INFO - Iter 4201/10000, Loss: 0.2159, Avg Adaptive LR: 3.294529
2025-03-18 21:29:57,671 - INFO - Iter 4211/10000, Loss: 0.2159, Avg Adaptive LR: 3.294527
2025-03-18 21:29:58,083 - INFO - Iter 4221/10000, Loss: 0.2158, Avg Adaptive LR: 3.294526
2025-03-18 21:29:58,487 - INFO - Iter 4231/10000, Loss: 0.2158, Avg Adaptive LR: 3.294524
2025-03-18 21:29:58,878 - INFO - Iter 4241/10000, Loss: 0.2158, Avg Adaptive LR: 3.294522
2025-03-18 21:29:59,284 - INFO - Iter 4251/10000, Loss: 0.2158, Avg Adaptive LR: 3.294521
2025-03-18 21:29:59,682 - INFO - Iter 4261/10000, Loss: 0.2157, Avg Adaptive LR: 3.294519
2025-03-18 21:30:00,080 - INFO - Iter 4271/10000, Loss: 0.2157, Avg Adaptive LR: 3.294517
2025-03-18 21:30:00,482 - INFO - Iter 4281/10000, Loss: 0.2157, Avg Adaptive LR: 3.294516
2025-03-18 21:30:00,877 - INFO - Iter 4291/10000, Loss: 0.2157, Avg Adaptive LR: 3.294514
2025-03-18 21:30:01,275 - INFO - Iter 4301/10000, Loss: 0.2156, Avg Adaptive LR: 3.294512
2025-03-18 21:30:01,667 - INFO - Iter 4311/10000, Loss: 0.2156, Avg Adaptive LR: 3.294511
2025-03-18 21:30:02,061 - INFO - Iter 4321/10000, Loss: 0.2156, Avg Adaptive LR: 3.294509
2025-03-18 21:30:02,460 - INFO - Iter 4331/10000, Loss: 0.2156, Avg Adaptive LR: 3.294508
2025-03-18 21:30:02,850 - INFO - Iter 4341/10000, Loss: 0.2156, Avg Adaptive LR: 3.294506
2025-03-18 21:30:03,251 - INFO - Iter 4351/10000, Loss: 0.2155, Avg Adaptive LR: 3.294504
2025-03-18 21:30:03,645 - INFO - Iter 4361/10000, Loss: 0.2155, Avg Adaptive LR: 3.294503
2025-03-18 21:30:04,036 - INFO - Iter 4371/10000, Loss: 0.2155, Avg Adaptive LR: 3.294501
2025-03-18 21:30:04,438 - INFO - Iter 4381/10000, Loss: 0.2155, Avg Adaptive LR: 3.294500
2025-03-18 21:30:04,833 - INFO - Iter 4391/10000, Loss: 0.2154, Avg Adaptive LR: 3.294498
2025-03-18 21:30:05,227 - INFO - Iter 4401/10000, Loss: 0.2154, Avg Adaptive LR: 3.294497
2025-03-18 21:30:05,626 - INFO - Iter 4411/10000, Loss: 0.2154, Avg Adaptive LR: 3.294495
2025-03-18 21:30:06,017 - INFO - Iter 4421/10000, Loss: 0.2154, Avg Adaptive LR: 3.294493
2025-03-18 21:30:06,418 - INFO - Iter 4431/10000, Loss: 0.2154, Avg Adaptive LR: 3.294492
2025-03-18 21:30:06,810 - INFO - Iter 4441/10000, Loss: 0.2153, Avg Adaptive LR: 3.294490
2025-03-18 21:30:07,206 - INFO - Iter 4451/10000, Loss: 0.2153, Avg Adaptive LR: 3.294489
2025-03-18 21:30:07,610 - INFO - Iter 4461/10000, Loss: 0.2153, Avg Adaptive LR: 3.294487
2025-03-18 21:30:08,034 - INFO - Iter 4471/10000, Loss: 0.2153, Avg Adaptive LR: 3.294486
2025-03-18 21:30:08,434 - INFO - Iter 4481/10000, Loss: 0.2152, Avg Adaptive LR: 3.294484
2025-03-18 21:30:08,832 - INFO - Iter 4491/10000, Loss: 0.2152, Avg Adaptive LR: 3.294483
```

```
2025-03-18 21:30:09,227 - INFO - Iter 4501/10000, Loss: 0.2152, Avg Adaptive LR: 3.294481
2025-03-18 21:30:09,626 - INFO - Iter 4511/10000, Loss: 0.2152, Avg Adaptive LR: 3.294480
2025-03-18 21:30:10,019 - INFO - Iter 4521/10000, Loss: 0.2152, Avg Adaptive LR: 3.294478
2025-03-18 21:30:10,428 - INFO - Iter 4531/10000, Loss: 0.2151, Avg Adaptive LR: 3.294477
2025-03-18 21:30:10,825 - INFO - Iter 4541/10000, Loss: 0.2151, Avg Adaptive LR: 3.294475
2025-03-18 21:30:11,221 - INFO - Iter 4551/10000, Loss: 0.2151, Avg Adaptive LR: 3.294474
2025-03-18 21:30:11,628 - INFO - Iter 4561/10000, Loss: 0.2151, Avg Adaptive LR: 3.294472
2025-03-18 21:30:12,019 - INFO - Iter 4571/10000, Loss: 0.2150, Avg Adaptive LR: 3.294471
2025-03-18 21:30:12,427 - INFO - Iter 4581/10000, Loss: 0.2150, Avg Adaptive LR: 3.294469
2025-03-18 21:30:12,820 - INFO - Iter 4591/10000, Loss: 0.2150, Avg Adaptive LR: 3.294468
2025-03-18 21:30:13,218 - INFO - Iter 4601/10000, Loss: 0.2150, Avg Adaptive LR: 3.294466
2025-03-18 21:30:13,616 - INFO - Iter 4611/10000, Loss: 0.2150, Avg Adaptive LR: 3.294465
2025-03-18 21:30:14,013 - INFO - Iter 4621/10000, Loss: 0.2149, Avg Adaptive LR: 3.294463
2025-03-18 21:30:14,419 - INFO - Iter 4631/10000, Loss: 0.2149, Avg Adaptive LR: 3.294462
2025-03-18 21:30:14,814 - INFO - Iter 4641/10000, Loss: 0.2149, Avg Adaptive LR: 3.294460
2025-03-18 21:30:15,204 - INFO - Iter 4651/10000, Loss: 0.2149, Avg Adaptive LR: 3.294459
2025-03-18 21:30:15,605 - INFO - Iter 4661/10000, Loss: 0.2149, Avg Adaptive LR: 3.294458
2025-03-18 21:30:15,999 - INFO - Iter 4671/10000, Loss: 0.2148, Avg Adaptive LR: 3.294456
2025-03-18 21:30:16,402 - INFO - Iter 4681/10000, Loss: 0.2148, Avg Adaptive LR: 3.294455
2025-03-18 21:30:16,802 - INFO - Iter 4691/10000, Loss: 0.2148, Avg Adaptive LR: 3.294453
2025-03-18 21:30:17,203 - INFO - Iter 4701/10000, Loss: 0.2148, Avg Adaptive LR: 3.294452
2025-03-18 21:30:17,607 - INFO - Iter 4711/10000, Loss: 0.2147, Avg Adaptive LR: 3.294450
2025-03-18 21:30:18,011 - INFO - Iter 4721/10000, Loss: 0.2147, Avg Adaptive LR: 3.294449
2025-03-18 21:30:18,430 - INFO - Iter 4731/10000, Loss: 0.2147, Avg Adaptive LR: 3.294448
2025-03-18 21:30:18,825 - INFO - Iter 4741/10000, Loss: 0.2147, Avg Adaptive LR: 3.294446
2025-03-18 21:30:19,219 - INFO - Iter 4751/10000, Loss: 0.2147, Avg Adaptive LR: 3.294445
2025-03-18 21:30:19,618 - INFO - Iter 4761/10000, Loss: 0.2146, Avg Adaptive LR: 3.294443
2025-03-18 21:30:20,013 - INFO - Iter 4771/10000, Loss: 0.2146, Avg Adaptive LR: 3.294442
2025-03-18 21:30:20,415 - INFO - Iter 4781/10000, Loss: 0.2146, Avg Adaptive LR: 3.294441
2025-03-18 21:30:20,811 - INFO - Iter 4791/10000, Loss: 0.2146, Avg Adaptive LR: 3.294439
2025-03-18 21:30:21,204 - INFO - Iter 4801/10000, Loss: 0.2146, Avg Adaptive LR: 3.294438
2025-03-18 21:30:21,611 - INFO - Iter 4811/10000, Loss: 0.2145, Avg Adaptive LR: 3.294436
2025-03-18 21:30:22,003 - INFO - Iter 4821/10000, Loss: 0.2145, Avg Adaptive LR: 3.294435
2025-03-18 21:30:22,415 - INFO - Iter 4831/10000, Loss: 0.2145, Avg Adaptive LR: 3.294434
2025-03-18 21:30:22,821 - INFO - Iter 4841/10000, Loss: 0.2145, Avg Adaptive LR: 3.294432
2025-03-18 21:30:23,215 - INFO - Iter 4851/10000, Loss: 0.2145, Avg Adaptive LR: 3.294431
2025-03-18 21:30:23,615 - INFO - Iter 4861/10000, Loss: 0.2144, Avg Adaptive LR: 3.294430
2025-03-18 21:30:24,019 - INFO - Iter 4871/10000, Loss: 0.2144, Avg Adaptive LR: 3.294428
2025-03-18 21:30:24,423 - INFO - Iter 4881/10000, Loss: 0.2144, Avg Adaptive LR: 3.294427
2025-03-18 21:30:24,814 - INFO - Iter 4891/10000, Loss: 0.2144, Avg Adaptive LR: 3.294426
2025-03-18 21:30:25,209 - INFO - Iter 4901/10000, Loss: 0.2144, Avg Adaptive LR: 3.294424
2025-03-18 21:30:25,609 - INFO - Iter 4911/10000, Loss: 0.2144, Avg Adaptive LR: 3.294423
2025-03-18 21:30:25,998 - INFO - Iter 4921/10000, Loss: 0.2143, Avg Adaptive LR: 3.294422
2025-03-18 21:30:26,401 - INFO - Iter 4931/10000, Loss: 0.2143, Avg Adaptive LR: 3.294420
2025-03-18 21:30:26,799 - INFO - Iter 4941/10000, Loss: 0.2143, Avg Adaptive LR: 3.294419
2025-03-18 21:30:27,210 - INFO - Iter 4951/10000, Loss: 0.2143, Avg Adaptive LR: 3.294418
2025-03-18 21:30:27,609 - INFO - Iter 4961/10000, Loss: 0.2143, Avg Adaptive LR: 3.294416
2025-03-18 21:30:28,008 - INFO - Iter 4971/10000, Loss: 0.2142, Avg Adaptive LR: 3.294415
2025-03-18 21:30:28,424 - INFO - Iter 4981/10000, Loss: 0.2142, Avg Adaptive LR: 3.294414
2025-03-18 21:30:28,821 - INFO - Iter 4991/10000, Loss: 0.2142, Avg Adaptive LR: 3.294412
2025-03-18 21:30:29,216 - INFO - Iter 5001/10000, Loss: 0.2142, Avg Adaptive LR: 3.294411
2025-03-18 21:30:29,616 - INFO - Iter 5011/10000, Loss: 0.2142, Avg Adaptive LR: 3.294410
2025-03-18 21:30:30,010 - INFO - Iter 5021/10000, Loss: 0.2141, Avg Adaptive LR: 3.294408
2025-03-18 21:30:30,409 - INFO - Iter 5031/10000, Loss: 0.2141, Avg Adaptive LR: 3.294407
2025-03-18 21:30:30,804 - INFO - Iter 5041/10000, Loss: 0.2141, Avg Adaptive LR: 3.294406
2025-03-18 21:30:31,197 - INFO - Iter 5051/10000, Loss: 0.2141, Avg Adaptive LR: 3.294405
2025-03-18 21:30:31,600 - INFO - Iter 5061/10000, Loss: 0.2141, Avg Adaptive LR: 3.294403
2025-03-18 21:30:31,997 - INFO - Iter 5071/10000, Loss: 0.2140, Avg Adaptive LR: 3.294402
2025-03-18 21:30:32,398 - INFO - Iter 5081/10000, Loss: 0.2140, Avg Adaptive LR: 3.294401
2025-03-18 21:30:32,794 - INFO - Iter 5091/10000, Loss: 0.2140, Avg Adaptive LR: 3.294399
2025-03-18 21:30:33,200 - INFO - Iter 5101/10000, Loss: 0.2140, Avg Adaptive LR: 3.294398
2025-03-18 21:30:33,603 - INFO - Iter 5111/10000, Loss: 0.2140, Avg Adaptive LR: 3.294397
2025-03-18 21:30:33,996 - INFO - Iter 5121/10000, Loss: 0.2140, Avg Adaptive LR: 3.294396
2025-03-18 21:30:34,395 - INFO - Iter 5131/10000, Loss: 0.2139, Avg Adaptive LR: 3.294394
2025-03-18 21:30:34,790 - INFO - Iter 5141/10000, Loss: 0.2139, Avg Adaptive LR: 3.294393
2025-03-18 21:30:35,184 - INFO - Iter 5151/10000, Loss: 0.2139, Avg Adaptive LR: 3.294392
2025-03-18 21:30:35,584 - INFO - Iter 5161/10000, Loss: 0.2139, Avg Adaptive LR: 3.294391
2025-03-18 21:30:35,975 - INFO - Iter 5171/10000, Loss: 0.2139, Avg Adaptive LR: 3.294389
2025-03-18 21:30:36,379 - INFO - Iter 5181/10000, Loss: 0.2138, Avg Adaptive LR: 3.294388
2025-03-18 21:30:36,775 - INFO - Iter 5191/10000, Loss: 0.2138, Avg Adaptive LR: 3.294387
2025-03-18 21:30:37,169 - INFO - Iter 5201/10000, Loss: 0.2138, Avg Adaptive LR: 3.294386
2025-03-18 21:30:37,583 - INFO - Iter 5211/10000, Loss: 0.2138, Avg Adaptive LR: 3.294384
```

```
2025-03-18 21:30:37,977 - INFO - Iter 5221/10000, Loss: 0.2138, Avg Adaptive LR: 3.294383
2025-03-18 21:30:38,391 - INFO - Iter 5231/10000, Loss: 0.2138, Avg Adaptive LR: 3.294382
2025-03-18 21:30:38,784 - INFO - Iter 5241/10000, Loss: 0.2137, Avg Adaptive LR: 3.294381
2025-03-18 21:30:39,176 - INFO - Iter 5251/10000, Loss: 0.2137, Avg Adaptive LR: 3.294380
2025-03-18 21:30:39,575 - INFO - Iter 5261/10000, Loss: 0.2137, Avg Adaptive LR: 3.294378
2025-03-18 21:30:39,968 - INFO - Iter 5271/10000, Loss: 0.2137, Avg Adaptive LR: 3.294377
2025-03-18 21:30:40,374 - INFO - Iter 5281/10000, Loss: 0.2137, Avg Adaptive LR: 3.294376
2025-03-18 21:30:40,769 - INFO - Iter 5291/10000, Loss: 0.2137, Avg Adaptive LR: 3.294375
2025-03-18 21:30:41,165 - INFO - Iter 5301/10000, Loss: 0.2136, Avg Adaptive LR: 3.294374
2025-03-18 21:30:41,568 - INFO - Iter 5311/10000, Loss: 0.2136, Avg Adaptive LR: 3.294372
2025-03-18 21:30:41,960 - INFO - Iter 5321/10000, Loss: 0.2136, Avg Adaptive LR: 3.294371
2025-03-18 21:30:42,365 - INFO - Iter 5331/10000, Loss: 0.2136, Avg Adaptive LR: 3.294370
2025-03-18 21:30:42,757 - INFO - Iter 5341/10000, Loss: 0.2136, Avg Adaptive LR: 3.294369
2025-03-18 21:30:43,151 - INFO - Iter 5351/10000, Loss: 0.2136, Avg Adaptive LR: 3.294368
2025-03-18 21:30:43,556 - INFO - Iter 5361/10000, Loss: 0.2135, Avg Adaptive LR: 3.294366
2025-03-18 21:30:43,950 - INFO - Iter 5371/10000, Loss: 0.2135, Avg Adaptive LR: 3.294365
2025-03-18 21:30:44,357 - INFO - Iter 5381/10000, Loss: 0.2135, Avg Adaptive LR: 3.294364
2025-03-18 21:30:44,751 - INFO - Iter 5391/10000, Loss: 0.2135, Avg Adaptive LR: 3.294363
2025-03-18 21:30:45,146 - INFO - Iter 5401/10000, Loss: 0.2135, Avg Adaptive LR: 3.294362
2025-03-18 21:30:45,549 - INFO - Iter 5411/10000, Loss: 0.2135, Avg Adaptive LR: 3.294361
2025-03-18 21:30:45,947 - INFO - Iter 5421/10000, Loss: 0.2134, Avg Adaptive LR: 3.294359
2025-03-18 21:30:46,350 - INFO - Iter 5431/10000, Loss: 0.2134, Avg Adaptive LR: 3.294358
2025-03-18 21:30:46,740 - INFO - Iter 5441/10000, Loss: 0.2134, Avg Adaptive LR: 3.294357
2025-03-18 21:30:47,132 - INFO - Iter 5451/10000, Loss: 0.2134, Avg Adaptive LR: 3.294356
2025-03-18 21:30:47,558 - INFO - Iter 5461/10000, Loss: 0.2134, Avg Adaptive LR: 3.294355
2025-03-18 21:30:47,953 - INFO - Iter 5471/10000, Loss: 0.2134, Avg Adaptive LR: 3.294354
2025-03-18 21:30:48,364 - INFO - Iter 5481/10000, Loss: 0.2133, Avg Adaptive LR: 3.294353
2025-03-18 21:30:48,774 - INFO - Iter 5491/10000, Loss: 0.2133, Avg Adaptive LR: 3.294351
2025-03-18 21:30:49,165 - INFO - Iter 5501/10000, Loss: 0.2133, Avg Adaptive LR: 3.294350
2025-03-18 21:30:49,568 - INFO - Iter 5511/10000, Loss: 0.2133, Avg Adaptive LR: 3.294349
2025-03-18 21:30:49,961 - INFO - Iter 5521/10000, Loss: 0.2133, Avg Adaptive LR: 3.294348
2025-03-18 21:30:50,366 - INFO - Iter 5531/10000, Loss: 0.2133, Avg Adaptive LR: 3.294347
2025-03-18 21:30:50,760 - INFO - Iter 5541/10000, Loss: 0.2132, Avg Adaptive LR: 3.294346
2025-03-18 21:30:51,148 - INFO - Iter 5551/10000, Loss: 0.2132, Avg Adaptive LR: 3.294345
2025-03-18 21:30:51,549 - INFO - Iter 5561/10000, Loss: 0.2132, Avg Adaptive LR: 3.294344
2025-03-18 21:30:51,948 - INFO - Iter 5571/10000, Loss: 0.2132, Avg Adaptive LR: 3.294342
2025-03-18 21:30:52,350 - INFO - Iter 5581/10000, Loss: 0.2132, Avg Adaptive LR: 3.294341
2025-03-18 21:30:52,744 - INFO - Iter 5591/10000, Loss: 0.2132, Avg Adaptive LR: 3.294340
2025-03-18 21:30:53,135 - INFO - Iter 5601/10000, Loss: 0.2131, Avg Adaptive LR: 3.294339
2025-03-18 21:30:53,540 - INFO - Iter 5611/10000, Loss: 0.2131, Avg Adaptive LR: 3.294338
2025-03-18 21:30:53,934 - INFO - Iter 5621/10000, Loss: 0.2131, Avg Adaptive LR: 3.294337
2025-03-18 21:30:54,338 - INFO - Iter 5631/10000, Loss: 0.2131, Avg Adaptive LR: 3.294336
2025-03-18 21:30:54,732 - INFO - Iter 5641/10000, Loss: 0.2131, Avg Adaptive LR: 3.294335
2025-03-18 21:30:55,127 - INFO - Iter 5651/10000, Loss: 0.2131, Avg Adaptive LR: 3.294334
2025-03-18 21:30:55,534 - INFO - Iter 5661/10000, Loss: 0.2130, Avg Adaptive LR: 3.294333
2025-03-18 21:30:55,929 - INFO - Iter 5671/10000, Loss: 0.2130, Avg Adaptive LR: 3.294331
2025-03-18 21:30:56,329 - INFO - Iter 5681/10000, Loss: 0.2130, Avg Adaptive LR: 3.294330
2025-03-18 21:30:56,728 - INFO - Iter 5691/10000, Loss: 0.2130, Avg Adaptive LR: 3.294329
2025-03-18 21:30:57,126 - INFO - Iter 5701/10000, Loss: 0.2130, Avg Adaptive LR: 3.294328
2025-03-18 21:30:57,541 - INFO - Iter 5711/10000, Loss: 0.2130, Avg Adaptive LR: 3.294327
2025-03-18 21:30:57,958 - INFO - Iter 5721/10000, Loss: 0.2130, Avg Adaptive LR: 3.294326
2025-03-18 21:30:58,367 - INFO - Iter 5731/10000, Loss: 0.2129, Avg Adaptive LR: 3.294325
2025-03-18 21:30:58,776 - INFO - Iter 5741/10000, Loss: 0.2129, Avg Adaptive LR: 3.294324
2025-03-18 21:30:59,170 - INFO - Iter 5751/10000, Loss: 0.2129, Avg Adaptive LR: 3.294323
2025-03-18 21:30:59,581 - INFO - Iter 5761/10000, Loss: 0.2129, Avg Adaptive LR: 3.294322
2025-03-18 21:30:59,983 - INFO - Iter 5771/10000, Loss: 0.2129, Avg Adaptive LR: 3.294321
2025-03-18 21:31:00,390 - INFO - Iter 5781/10000, Loss: 0.2129, Avg Adaptive LR: 3.294320
2025-03-18 21:31:00,790 - INFO - Iter 5791/10000, Loss: 0.2128, Avg Adaptive LR: 3.294319
2025-03-18 21:31:01,182 - INFO - Iter 5801/10000, Loss: 0.2128, Avg Adaptive LR: 3.294318
2025-03-18 21:31:01,580 - INFO - Iter 5811/10000, Loss: 0.2128, Avg Adaptive LR: 3.294317
2025-03-18 21:31:01,968 - INFO - Iter 5821/10000, Loss: 0.2128, Avg Adaptive LR: 3.294316
2025-03-18 21:31:02,380 - INFO - Iter 5831/10000, Loss: 0.2128, Avg Adaptive LR: 3.294314
2025-03-18 21:31:02,771 - INFO - Iter 5841/10000, Loss: 0.2128, Avg Adaptive LR: 3.294313
2025-03-18 21:31:03,164 - INFO - Iter 5851/10000, Loss: 0.2128, Avg Adaptive LR: 3.294312
2025-03-18 21:31:03,566 - INFO - Iter 5861/10000, Loss: 0.2127, Avg Adaptive LR: 3.294311
2025-03-18 21:31:03,965 - INFO - Iter 5871/10000, Loss: 0.2127, Avg Adaptive LR: 3.294310
2025-03-18 21:31:04,366 - INFO - Iter 5881/10000, Loss: 0.2127, Avg Adaptive LR: 3.294309
2025-03-18 21:31:04,762 - INFO - Iter 5891/10000, Loss: 0.2127, Avg Adaptive LR: 3.294308
2025-03-18 21:31:05,154 - INFO - Iter 5901/10000, Loss: 0.2127, Avg Adaptive LR: 3.294307
2025-03-18 21:31:05,552 - INFO - Iter 5911/10000, Loss: 0.2127, Avg Adaptive LR: 3.294306
2025-03-18 21:31:05,945 - INFO - Iter 5921/10000, Loss: 0.2126, Avg Adaptive LR: 3.294305
2025-03-18 21:31:06,350 - INFO - Iter 5931/10000, Loss: 0.2126, Avg Adaptive LR: 3.294304
```

```
2025-03-18 21:31:06,742 - INFO - Iter 5941/10000, Loss: 0.2126, Avg Adaptive LR: 3.294303
2025-03-18 21:31:07,136 - INFO - Iter 5951/10000, Loss: 0.2126, Avg Adaptive LR: 3.294302
2025-03-18 21:31:07,541 - INFO - Iter 5961/10000, Loss: 0.2126, Avg Adaptive LR: 3.294301
2025-03-18 21:31:07,934 - INFO - Iter 5971/10000, Loss: 0.2126, Avg Adaptive LR: 3.294300
2025-03-18 21:31:08,349 - INFO - Iter 5981/10000, Loss: 0.2126, Avg Adaptive LR: 3.294299
2025-03-18 21:31:08,745 - INFO - Iter 5991/10000, Loss: 0.2125, Avg Adaptive LR: 3.294298
2025-03-18 21:31:09,147 - INFO - Iter 6001/10000, Loss: 0.2125, Avg Adaptive LR: 3.294297
2025-03-18 21:31:09,547 - INFO - Iter 6011/10000, Loss: 0.2125, Avg Adaptive LR: 3.294296
2025-03-18 21:31:09,941 - INFO - Iter 6021/10000, Loss: 0.2125, Avg Adaptive LR: 3.294295
2025-03-18 21:31:10,340 - INFO - Iter 6031/10000, Loss: 0.2125, Avg Adaptive LR: 3.294294
2025-03-18 21:31:10,730 - INFO - Iter 6041/10000, Loss: 0.2125, Avg Adaptive LR: 3.294293
2025-03-18 21:31:11,125 - INFO - Iter 6051/10000, Loss: 0.2125, Avg Adaptive LR: 3.294292
2025-03-18 21:31:11,531 - INFO - Iter 6061/10000, Loss: 0.2124, Avg Adaptive LR: 3.294291
2025-03-18 21:31:11,927 - INFO - Iter 6071/10000, Loss: 0.2124, Avg Adaptive LR: 3.294290
2025-03-18 21:31:12,329 - INFO - Iter 6081/10000, Loss: 0.2124, Avg Adaptive LR: 3.294289
2025-03-18 21:31:12,719 - INFO - Iter 6091/10000, Loss: 0.2124, Avg Adaptive LR: 3.294288
2025-03-18 21:31:13,115 - INFO - Iter 6101/10000, Loss: 0.2124, Avg Adaptive LR: 3.294287
2025-03-18 21:31:13,518 - INFO - Iter 6111/10000, Loss: 0.2124, Avg Adaptive LR: 3.294286
2025-03-18 21:31:13,921 - INFO - Iter 6121/10000, Loss: 0.2124, Avg Adaptive LR: 3.294285
2025-03-18 21:31:14,321 - INFO - Iter 6131/10000, Loss: 0.2123, Avg Adaptive LR: 3.294284
2025-03-18 21:31:14,713 - INFO - Iter 6141/10000, Loss: 0.2123, Avg Adaptive LR: 3.294283
2025-03-18 21:31:15,109 - INFO - Iter 6151/10000, Loss: 0.2123, Avg Adaptive LR: 3.294282
2025-03-18 21:31:15,512 - INFO - Iter 6161/10000, Loss: 0.2123, Avg Adaptive LR: 3.294282
2025-03-18 21:31:15,910 - INFO - Iter 6171/10000, Loss: 0.2123, Avg Adaptive LR: 3.294281
2025-03-18 21:31:16,316 - INFO - Iter 6181/10000, Loss: 0.2123, Avg Adaptive LR: 3.294280
2025-03-18 21:31:16,711 - INFO - Iter 6191/10000, Loss: 0.2123, Avg Adaptive LR: 3.294279
2025-03-18 21:31:17,103 - INFO - Iter 6201/10000, Loss: 0.2123, Avg Adaptive LR: 3.294278
2025-03-18 21:31:17,509 - INFO - Iter 6211/10000, Loss: 0.2122, Avg Adaptive LR: 3.294277
2025-03-18 21:31:17,909 - INFO - Iter 6221/10000, Loss: 0.2122, Avg Adaptive LR: 3.294276
2025-03-18 21:31:18,314 - INFO - Iter 6231/10000, Loss: 0.2122, Avg Adaptive LR: 3.294275
2025-03-18 21:31:18,723 - INFO - Iter 6241/10000, Loss: 0.2122, Avg Adaptive LR: 3.294274
2025-03-18 21:31:19,131 - INFO - Iter 6251/10000, Loss: 0.2122, Avg Adaptive LR: 3.294273
2025-03-18 21:31:19,537 - INFO - Iter 6261/10000, Loss: 0.2122, Avg Adaptive LR: 3.294272
2025-03-18 21:31:19,930 - INFO - Iter 6271/10000, Loss: 0.2122, Avg Adaptive LR: 3.294271
2025-03-18 21:31:20,334 - INFO - Iter 6281/10000, Loss: 0.2121, Avg Adaptive LR: 3.294270
2025-03-18 21:31:20,730 - INFO - Iter 6291/10000, Loss: 0.2121, Avg Adaptive LR: 3.294269
2025-03-18 21:31:21,130 - INFO - Iter 6301/10000, Loss: 0.2121, Avg Adaptive LR: 3.294268
2025-03-18 21:31:21,531 - INFO - Iter 6311/10000, Loss: 0.2121, Avg Adaptive LR: 3.294267
2025-03-18 21:31:21,927 - INFO - Iter 6321/10000, Loss: 0.2121, Avg Adaptive LR: 3.294266
2025-03-18 21:31:22,334 - INFO - Iter 6331/10000, Loss: 0.2121, Avg Adaptive LR: 3.294266
2025-03-18 21:31:22,731 - INFO - Iter 6341/10000, Loss: 0.2121, Avg Adaptive LR: 3.294265
2025-03-18 21:31:23,128 - INFO - Iter 6351/10000, Loss: 0.2120, Avg Adaptive LR: 3.294264
2025-03-18 21:31:23,533 - INFO - Iter 6361/10000, Loss: 0.2120, Avg Adaptive LR: 3.294263
2025-03-18 21:31:23,933 - INFO - Iter 6371/10000, Loss: 0.2120, Avg Adaptive LR: 3.294262
2025-03-18 21:31:24,338 - INFO - Iter 6381/10000, Loss: 0.2120, Avg Adaptive LR: 3.294261
2025-03-18 21:31:24,734 - INFO - Iter 6391/10000, Loss: 0.2120, Avg Adaptive LR: 3.294260
2025-03-18 21:31:25,134 - INFO - Iter 6401/10000, Loss: 0.2120, Avg Adaptive LR: 3.294259
2025-03-18 21:31:25,533 - INFO - Iter 6411/10000, Loss: 0.2120, Avg Adaptive LR: 3.294258
2025-03-18 21:31:25,929 - INFO - Iter 6421/10000, Loss: 0.2120, Avg Adaptive LR: 3.294257
2025-03-18 21:31:26,330 - INFO - Iter 6431/10000, Loss: 0.2119, Avg Adaptive LR: 3.294256
2025-03-18 21:31:26,727 - INFO - Iter 6441/10000, Loss: 0.2119, Avg Adaptive LR: 3.294256
2025-03-18 21:31:27,117 - INFO - Iter 6451/10000, Loss: 0.2119, Avg Adaptive LR: 3.294255
2025-03-18 21:31:27,515 - INFO - Iter 6461/10000, Loss: 0.2119, Avg Adaptive LR: 3.294254
2025-03-18 21:31:27,908 - INFO - Iter 6471/10000, Loss: 0.2119, Avg Adaptive LR: 3.294253
2025-03-18 21:31:28,312 - INFO - Iter 6481/10000, Loss: 0.2119, Avg Adaptive LR: 3.294252
2025-03-18 21:31:28,718 - INFO - Iter 6491/10000, Loss: 0.2119, Avg Adaptive LR: 3.294251
2025-03-18 21:31:29,133 - INFO - Iter 6501/10000, Loss: 0.2119, Avg Adaptive LR: 3.294250
2025-03-18 21:31:29,533 - INFO - Iter 6511/10000, Loss: 0.2118, Avg Adaptive LR: 3.294249
2025-03-18 21:31:29,928 - INFO - Iter 6521/10000, Loss: 0.2118, Avg Adaptive LR: 3.294248
2025-03-18 21:31:30,325 - INFO - Iter 6531/10000, Loss: 0.2118, Avg Adaptive LR: 3.294248
2025-03-18 21:31:30,724 - INFO - Iter 6541/10000, Loss: 0.2118, Avg Adaptive LR: 3.294247
2025-03-18 21:31:31,118 - INFO - Iter 6551/10000, Loss: 0.2118, Avg Adaptive LR: 3.294246
2025-03-18 21:31:31,520 - INFO - Iter 6561/10000, Loss: 0.2118, Avg Adaptive LR: 3.294245
2025-03-18 21:31:31,918 - INFO - Iter 6571/10000, Loss: 0.2118, Avg Adaptive LR: 3.294244
2025-03-18 21:31:32,320 - INFO - Iter 6581/10000, Loss: 0.2118, Avg Adaptive LR: 3.294243
2025-03-18 21:31:32,714 - INFO - Iter 6591/10000, Loss: 0.2117, Avg Adaptive LR: 3.294242
2025-03-18 21:31:33,110 - INFO - Iter 6601/10000, Loss: 0.2117, Avg Adaptive LR: 3.294241
2025-03-18 21:31:33,510 - INFO - Iter 6611/10000, Loss: 0.2117, Avg Adaptive LR: 3.294241
2025-03-18 21:31:33,908 - INFO - Iter 6621/10000, Loss: 0.2117, Avg Adaptive LR: 3.294240
2025-03-18 21:31:34,309 - INFO - Iter 6631/10000, Loss: 0.2117, Avg Adaptive LR: 3.294239
2025-03-18 21:31:34,704 - INFO - Iter 6641/10000, Loss: 0.2117, Avg Adaptive LR: 3.294238
2025-03-18 21:31:35,098 - INFO - Iter 6651/10000, Loss: 0.2117, Avg Adaptive LR: 3.294237
```

```
2025-03-18 21:31:35,500 - INFO - Iter 6661/10000, Loss: 0.2117, Avg Adaptive LR: 3.294236
2025-03-18 21:31:35,893 - INFO - Iter 6671/10000, Loss: 0.2116, Avg Adaptive LR: 3.294235
2025-03-18 21:31:36,297 - INFO - Iter 6681/10000, Loss: 0.2116, Avg Adaptive LR: 3.294235
2025-03-18 21:31:36,694 - INFO - Iter 6691/10000, Loss: 0.2116, Avg Adaptive LR: 3.294234
2025-03-18 21:31:37,085 - INFO - Iter 6701/10000, Loss: 0.2116, Avg Adaptive LR: 3.294233
2025-03-18 21:31:37,489 - INFO - Iter 6711/10000, Loss: 0.2116, Avg Adaptive LR: 3.294232
2025-03-18 21:31:37,883 - INFO - Iter 6721/10000, Loss: 0.2116, Avg Adaptive LR: 3.294231
2025-03-18 21:31:38,287 - INFO - Iter 6731/10000, Loss: 0.2116, Avg Adaptive LR: 3.294230
2025-03-18 21:31:38,680 - INFO - Iter 6741/10000, Loss: 0.2116, Avg Adaptive LR: 3.294229
2025-03-18 21:31:39,070 - INFO - Iter 6751/10000, Loss: 0.2115, Avg Adaptive LR: 3.294229
2025-03-18 21:31:39,496 - INFO - Iter 6761/10000, Loss: 0.2115, Avg Adaptive LR: 3.294228
2025-03-18 21:31:39,897 - INFO - Iter 6771/10000, Loss: 0.2115, Avg Adaptive LR: 3.294227
2025-03-18 21:31:40,295 - INFO - Iter 6781/10000, Loss: 0.2115, Avg Adaptive LR: 3.294226
2025-03-18 21:31:40,690 - INFO - Iter 6791/10000, Loss: 0.2115, Avg Adaptive LR: 3.294225
2025-03-18 21:31:41,080 - INFO - Iter 6801/10000, Loss: 0.2115, Avg Adaptive LR: 3.294224
2025-03-18 21:31:41,486 - INFO - Iter 6811/10000, Loss: 0.2115, Avg Adaptive LR: 3.294224
2025-03-18 21:31:41,889 - INFO - Iter 6821/10000, Loss: 0.2115, Avg Adaptive LR: 3.294223
2025-03-18 21:31:42,292 - INFO - Iter 6831/10000, Loss: 0.2114, Avg Adaptive LR: 3.294222
2025-03-18 21:31:42,686 - INFO - Iter 6841/10000, Loss: 0.2114, Avg Adaptive LR: 3.294221
2025-03-18 21:31:43,077 - INFO - Iter 6851/10000, Loss: 0.2114, Avg Adaptive LR: 3.294220
2025-03-18 21:31:43,474 - INFO - Iter 6861/10000, Loss: 0.2114, Avg Adaptive LR: 3.294219
2025-03-18 21:31:43,869 - INFO - Iter 6871/10000, Loss: 0.2114, Avg Adaptive LR: 3.294219
2025-03-18 21:31:44,276 - INFO - Iter 6881/10000, Loss: 0.2114, Avg Adaptive LR: 3.294218
2025-03-18 21:31:44,673 - INFO - Iter 6891/10000, Loss: 0.2114, Avg Adaptive LR: 3.294217
2025-03-18 21:31:45,069 - INFO - Iter 6901/10000, Loss: 0.2114, Avg Adaptive LR: 3.294216
2025-03-18 21:31:45,478 - INFO - Iter 6911/10000, Loss: 0.2113, Avg Adaptive LR: 3.294215
2025-03-18 21:31:45,879 - INFO - Iter 6921/10000, Loss: 0.2113, Avg Adaptive LR: 3.294215
2025-03-18 21:31:46,290 - INFO - Iter 6931/10000, Loss: 0.2113, Avg Adaptive LR: 3.294214
2025-03-18 21:31:46,681 - INFO - Iter 6941/10000, Loss: 0.2113, Avg Adaptive LR: 3.294213
2025-03-18 21:31:47,074 - INFO - Iter 6951/10000, Loss: 0.2113, Avg Adaptive LR: 3.294212
2025-03-18 21:31:47,476 - INFO - Iter 6961/10000, Loss: 0.2113, Avg Adaptive LR: 3.294211
2025-03-18 21:31:47,872 - INFO - Iter 6971/10000, Loss: 0.2113, Avg Adaptive LR: 3.294211
2025-03-18 21:31:48,276 - INFO - Iter 6981/10000, Loss: 0.2113, Avg Adaptive LR: 3.294210
2025-03-18 21:31:48,675 - INFO - Iter 6991/10000, Loss: 0.2113, Avg Adaptive LR: 3.294209
2025-03-18 21:31:49,067 - INFO - Iter 7001/10000, Loss: 0.2112, Avg Adaptive LR: 3.294208
2025-03-18 21:31:49,475 - INFO - Iter 7011/10000, Loss: 0.2112, Avg Adaptive LR: 3.294207
2025-03-18 21:31:49,889 - INFO - Iter 7021/10000, Loss: 0.2112, Avg Adaptive LR: 3.294207
2025-03-18 21:31:50,298 - INFO - Iter 7031/10000, Loss: 0.2112, Avg Adaptive LR: 3.294206
2025-03-18 21:31:50,691 - INFO - Iter 7041/10000, Loss: 0.2112, Avg Adaptive LR: 3.294205
2025-03-18 21:31:51,085 - INFO - Iter 7051/10000, Loss: 0.2112, Avg Adaptive LR: 3.294204
2025-03-18 21:31:51,491 - INFO - Iter 7061/10000, Loss: 0.2112, Avg Adaptive LR: 3.294203
2025-03-18 21:31:51,886 - INFO - Iter 7071/10000, Loss: 0.2112, Avg Adaptive LR: 3.294203
2025-03-18 21:31:52,288 - INFO - Iter 7081/10000, Loss: 0.2112, Avg Adaptive LR: 3.294202
2025-03-18 21:31:52,680 - INFO - Iter 7091/10000, Loss: 0.2111, Avg Adaptive LR: 3.294201
2025-03-18 21:31:53,078 - INFO - Iter 7101/10000, Loss: 0.2111, Avg Adaptive LR: 3.294200
2025-03-18 21:31:53,492 - INFO - Iter 7111/10000, Loss: 0.2111, Avg Adaptive LR: 3.294200
2025-03-18 21:31:53,885 - INFO - Iter 7121/10000, Loss: 0.2111, Avg Adaptive LR: 3.294199
2025-03-18 21:31:54,289 - INFO - Iter 7131/10000, Loss: 0.2111, Avg Adaptive LR: 3.294198
2025-03-18 21:31:54,692 - INFO - Iter 7141/10000, Loss: 0.2111, Avg Adaptive LR: 3.294197
2025-03-18 21:31:55,088 - INFO - Iter 7151/10000, Loss: 0.2111, Avg Adaptive LR: 3.294196
2025-03-18 21:31:55,490 - INFO - Iter 7161/10000, Loss: 0.2111, Avg Adaptive LR: 3.294196
2025-03-18 21:31:55,879 - INFO - Iter 7171/10000, Loss: 0.2111, Avg Adaptive LR: 3.294195
2025-03-18 21:31:56,287 - INFO - Iter 7181/10000, Loss: 0.2110, Avg Adaptive LR: 3.294194
2025-03-18 21:31:56,683 - INFO - Iter 7191/10000, Loss: 0.2110, Avg Adaptive LR: 3.294193
2025-03-18 21:31:57,077 - INFO - Iter 7201/10000, Loss: 0.2110, Avg Adaptive LR: 3.294193
2025-03-18 21:31:57,483 - INFO - Iter 7211/10000, Loss: 0.2110, Avg Adaptive LR: 3.294192
2025-03-18 21:31:57,881 - INFO - Iter 7221/10000, Loss: 0.2110, Avg Adaptive LR: 3.294191
2025-03-18 21:31:58,289 - INFO - Iter 7231/10000, Loss: 0.2110, Avg Adaptive LR: 3.294190
2025-03-18 21:31:58,695 - INFO - Iter 7241/10000, Loss: 0.2110, Avg Adaptive LR: 3.294190
2025-03-18 21:31:59,100 - INFO - Iter 7251/10000, Loss: 0.2110, Avg Adaptive LR: 3.294189
2025-03-18 21:31:59,500 - INFO - Iter 7261/10000, Loss: 0.2110, Avg Adaptive LR: 3.294188
2025-03-18 21:31:59,930 - INFO - Iter 7271/10000, Loss: 0.2109, Avg Adaptive LR: 3.294187
2025-03-18 21:32:00,341 - INFO - Iter 7281/10000, Loss: 0.2109, Avg Adaptive LR: 3.294187
2025-03-18 21:32:00,734 - INFO - Iter 7291/10000, Loss: 0.2109, Avg Adaptive LR: 3.294186
2025-03-18 21:32:01,128 - INFO - Iter 7301/10000, Loss: 0.2109, Avg Adaptive LR: 3.294185
2025-03-18 21:32:01,533 - INFO - Iter 7311/10000, Loss: 0.2109, Avg Adaptive LR: 3.294184
2025-03-18 21:32:01,928 - INFO - Iter 7321/10000, Loss: 0.2109, Avg Adaptive LR: 3.294184
2025-03-18 21:32:02,328 - INFO - Iter 7331/10000, Loss: 0.2109, Avg Adaptive LR: 3.294183
2025-03-18 21:32:02,728 - INFO - Iter 7341/10000, Loss: 0.2109, Avg Adaptive LR: 3.294182
2025-03-18 21:32:03,137 - INFO - Iter 7351/10000, Loss: 0.2109, Avg Adaptive LR: 3.294181
2025-03-18 21:32:03,540 - INFO - Iter 7361/10000, Loss: 0.2108, Avg Adaptive LR: 3.294181
2025-03-18 21:32:03,937 - INFO - Iter 7371/10000, Loss: 0.2108, Avg Adaptive LR: 3.294180
```

```
2025-03-18 21:32:04,356 - INFO - Iter 7381/10000, Loss: 0.2108, Avg Adaptive LR: 3.294179
2025-03-18 21:32:04,747 - INFO - Iter 7391/10000, Loss: 0.2108, Avg Adaptive LR: 3.294178
2025-03-18 21:32:05,143 - INFO - Iter 7401/10000, Loss: 0.2108, Avg Adaptive LR: 3.294178
2025-03-18 21:32:05,545 - INFO - Iter 7411/10000, Loss: 0.2108, Avg Adaptive LR: 3.294177
2025-03-18 21:32:05,944 - INFO - Iter 7421/10000, Loss: 0.2108, Avg Adaptive LR: 3.294176
2025-03-18 21:32:06,361 - INFO - Iter 7431/10000, Loss: 0.2108, Avg Adaptive LR: 3.294176
2025-03-18 21:32:06,761 - INFO - Iter 7441/10000, Loss: 0.2108, Avg Adaptive LR: 3.294175
2025-03-18 21:32:07,157 - INFO - Iter 7451/10000, Loss: 0.2108, Avg Adaptive LR: 3.294174
2025-03-18 21:32:07,566 - INFO - Iter 7461/10000, Loss: 0.2107, Avg Adaptive LR: 3.294173
2025-03-18 21:32:07,968 - INFO - Iter 7471/10000, Loss: 0.2107, Avg Adaptive LR: 3.294173
2025-03-18 21:32:08,373 - INFO - Iter 7481/10000, Loss: 0.2107, Avg Adaptive LR: 3.294172
2025-03-18 21:32:08,795 - INFO - Iter 7491/10000, Loss: 0.2107, Avg Adaptive LR: 3.294171
2025-03-18 21:32:09,200 - INFO - Iter 7501/10000, Loss: 0.2107, Avg Adaptive LR: 3.294171
2025-03-18 21:32:09,604 - INFO - Iter 7511/10000, Loss: 0.2107, Avg Adaptive LR: 3.294170
2025-03-18 21:32:10,014 - INFO - Iter 7521/10000, Loss: 0.2107, Avg Adaptive LR: 3.294169
2025-03-18 21:32:10,427 - INFO - Iter 7531/10000, Loss: 0.2107, Avg Adaptive LR: 3.294168
2025-03-18 21:32:10,826 - INFO - Iter 7541/10000, Loss: 0.2107, Avg Adaptive LR: 3.294168
2025-03-18 21:32:11,223 - INFO - Iter 7551/10000, Loss: 0.2106, Avg Adaptive LR: 3.294167
2025-03-18 21:32:11,637 - INFO - Iter 7561/10000, Loss: 0.2106, Avg Adaptive LR: 3.294166
2025-03-18 21:32:12,035 - INFO - Iter 7571/10000, Loss: 0.2106, Avg Adaptive LR: 3.294166
2025-03-18 21:32:12,447 - INFO - Iter 7581/10000, Loss: 0.2106, Avg Adaptive LR: 3.294165
2025-03-18 21:32:12,844 - INFO - Iter 7591/10000, Loss: 0.2106, Avg Adaptive LR: 3.294164
2025-03-18 21:32:13,236 - INFO - Iter 7601/10000, Loss: 0.2106, Avg Adaptive LR: 3.294163
2025-03-18 21:32:13,638 - INFO - Iter 7611/10000, Loss: 0.2106, Avg Adaptive LR: 3.294163
2025-03-18 21:32:14,036 - INFO - Iter 7621/10000, Loss: 0.2106, Avg Adaptive LR: 3.294162
2025-03-18 21:32:14,441 - INFO - Iter 7631/10000, Loss: 0.2106, Avg Adaptive LR: 3.294161
2025-03-18 21:32:14,835 - INFO - Iter 7641/10000, Loss: 0.2106, Avg Adaptive LR: 3.294161
2025-03-18 21:32:15,232 - INFO - Iter 7651/10000, Loss: 0.2105, Avg Adaptive LR: 3.294160
2025-03-18 21:32:15,630 - INFO - Iter 7661/10000, Loss: 0.2105, Avg Adaptive LR: 3.294159
2025-03-18 21:32:16,023 - INFO - Iter 7671/10000, Loss: 0.2105, Avg Adaptive LR: 3.294159
2025-03-18 21:32:16,429 - INFO - Iter 7681/10000, Loss: 0.2105, Avg Adaptive LR: 3.294158
2025-03-18 21:32:16,621 - INFO - Early stopping triggered at iteration 7686 with training
loss 0.210512
2025-03-18 21:32:16,622 - INFO - SoftmaxRegression training completed in 307.98 seconds.
2025-03-18 21:32:16,622 - INFO - --- Linear Regression LR=0.1/Iter=10000 ---
2025-03-18 21:32:19,793 - INFO - Iter 100/10000, Loss: 0.8123, Gradient Norm: 17.2331, Av
g Adaptive LR: 1.3966073375164807
2025-03-18 21:32:22,997 - INFO - Iter 200/10000, Loss: 0.4286, Gradient Norm: 12.2269, Av
g Adaptive LR: 0.9913722486345247
2025-03-18 21:32:26,248 - INFO - Iter 300/10000, Loss: 0.2906, Gradient Norm: 9.8200, Avg
Adaptive LR: 0.810855690898011
2025-03-18 21:32:29,526 - INFO - Iter 400/10000, Loss: 0.2200, Gradient Norm: 8.3230, Avg
Adaptive LR: 0.7029677295511769
2025-03-18 21:32:32,810 - INFO - Iter 500/10000, Loss: 0.1793, Gradient Norm: 7.3236, Avg
Adaptive LR: 0.6292007549960658
2025-03-18 21:32:36,099 - INFO - Iter 600/10000, Loss: 0.1520, Gradient Norm: 6.5704, Avg
Adaptive LR: 0.5746837799952516
2025-03-18 21:32:39,377 - INFO - Iter 700/10000, Loss: 0.1330, Gradient Norm: 5.9904, Avg
Adaptive LR: 0.532289891484879
2025-03-18 21:32:42,684 - INFO - Iter 800/10000, Loss: 0.1182, Gradient Norm: 5.4928, Avg
Adaptive LR: 0.4980731501312876
2025-03-18 21:32:46,028 - INFO - Iter 900/10000, Loss: 0.1077, Gradient Norm: 5.1159, Avg
Adaptive LR: 0.4697202013045508
2025-03-18 21:32:49,378 - INFO - Iter 1000/10000, Loss: 0.1009, Gradient Norm: 4.8531, Av
g Adaptive LR: 0.44572949233599296
2025-03-18 21:32:52,708 - INFO - Iter 1100/10000, Loss: 0.0930, Gradient Norm: 4.5317, Av
g Adaptive LR: 0.42506690371555017
2025-03-18 21:32:56,089 - INFO - Iter 1200/10000, Loss: 0.0851, Gradient Norm: 4.1839, Av
g Adaptive LR: 0.40703644929455796
2025-03-18 21:32:59,506 - INFO - Iter 1300/10000, Loss: 0.0804, Gradient Norm: 3.9600, Av
g Adaptive LR: 0.39113766904744734
2025-03-18 21:33:02,872 - INFO - Iter 1400/10000, Loss: 0.0750, Gradient Norm: 3.6925, Av
g Adaptive LR: 0.3769463937595444
2025-03-18 21:33:06,230 - INFO - Iter 1500/10000, Loss: 0.0706, Gradient Norm: 3.4557, Av
g Adaptive LR: 0.3642246466931996
2025-03-18 21:33:09,578 - INFO - Iter 1600/10000, Loss: 0.0663, Gradient Norm: 3.2042, Av
g Adaptive LR: 0.3526928071902571
2025-03-18 21:33:12,889 - INFO - Iter 1700/10000, Loss: 0.0634, Gradient Norm: 3.0285, Av
g Adaptive LR: 0.3421815082687663
2025-03-18 21:33:16,193 - INFO - Iter 1800/10000, Loss: 0.0611, Gradient Norm: 2.8820, Av
g Adaptive LR: 0.33256543370813635
2025-03-18 21:33:19,538 - INFO - Iter 1900/10000, Loss: 0.0592, Gradient Norm: 2.7519, Av
```

g Adaptive LR: 0.3237335626034096
2025-03-18 21:33:22,861 - INFO - Iter 2000/10000, Loss: 0.0580, Gradient Norm: 2.6673, Av
g Adaptive LR: 0.31556449240432066
2025-03-18 21:33:26,191 - INFO - Iter 2100/10000, Loss: 0.0567, Gradient Norm: 2.5695, Av
g Adaptive LR: 0.3079672287593788
2025-03-18 21:33:29,523 - INFO - Iter 2200/10000, Loss: 0.0556, Gradient Norm: 2.4883, Av
g Adaptive LR: 0.30091517138626694
2025-03-18 21:33:32,823 - INFO - Iter 2300/10000, Loss: 0.0546, Gradient Norm: 2.4127, Av
g Adaptive LR: 0.2943190170649717
2025-03-18 21:33:36,154 - INFO - Iter 2400/10000, Loss: 0.0538, Gradient Norm: 2.3548, Av
g Adaptive LR: 0.28813571220757567
2025-03-18 21:33:39,492 - INFO - Iter 2500/10000, Loss: 0.0528, Gradient Norm: 2.2707, Av
g Adaptive LR: 0.2823327677541542
2025-03-18 21:33:42,924 - INFO - Iter 2600/10000, Loss: 0.0524, Gradient Norm: 2.2352, Av
g Adaptive LR: 0.2768640557855329
2025-03-18 21:33:46,171 - INFO - Iter 2700/10000, Loss: 0.0516, Gradient Norm: 2.1668, Av
g Adaptive LR: 0.27170204740226594
2025-03-18 21:33:49,457 - INFO - Iter 2800/10000, Loss: 0.0509, Gradient Norm: 2.1088, Av
g Adaptive LR: 0.26681918951157113
2025-03-18 21:33:52,766 - INFO - Iter 2900/10000, Loss: 0.0508, Gradient Norm: 2.1008, Av
g Adaptive LR: 0.2621929033457426
2025-03-18 21:33:56,053 - INFO - Iter 3000/10000, Loss: 0.0498, Gradient Norm: 2.0099, Av
g Adaptive LR: 0.25779690709845
2025-03-18 21:33:59,394 - INFO - Iter 3100/10000, Loss: 0.0497, Gradient Norm: 1.9966, Av
g Adaptive LR: 0.25361489041229823
2025-03-18 21:34:02,729 - INFO - Iter 3200/10000, Loss: 0.0494, Gradient Norm: 1.9723, Av
g Adaptive LR: 0.24963232042995198
2025-03-18 21:34:06,024 - INFO - Iter 3300/10000, Loss: 0.0487, Gradient Norm: 1.9007, Av
g Adaptive LR: 0.245830706816637
2025-03-18 21:34:09,358 - INFO - Iter 3400/10000, Loss: 0.0487, Gradient Norm: 1.9024, Av
g Adaptive LR: 0.2421966453263162
2025-03-18 21:34:12,744 - INFO - Iter 3500/10000, Loss: 0.0483, Gradient Norm: 1.8632, Av
g Adaptive LR: 0.23872124796067673
2025-03-18 21:34:16,075 - INFO - Iter 3600/10000, Loss: 0.0478, Gradient Norm: 1.8157, Av
g Adaptive LR: 0.23538992538796852
2025-03-18 21:34:19,393 - INFO - Iter 3700/10000, Loss: 0.0478, Gradient Norm: 1.8187, Av
g Adaptive LR: 0.23219509484090406
2025-03-18 21:34:22,721 - INFO - Iter 3800/10000, Loss: 0.0474, Gradient Norm: 1.7714, Av
g Adaptive LR: 0.22912728699350474
2025-03-18 21:34:26,007 - INFO - Iter 3900/10000, Loss: 0.0471, Gradient Norm: 1.7408, Av
g Adaptive LR: 0.22617682587593463
2025-03-18 21:34:29,343 - INFO - Iter 4000/10000, Loss: 0.0471, Gradient Norm: 1.7438, Av
g Adaptive LR: 0.22333913965685556
2025-03-18 21:34:32,679 - INFO - Iter 4100/10000, Loss: 0.0467, Gradient Norm: 1.6967, Av
g Adaptive LR: 0.22060479475383526
2025-03-18 21:34:36,006 - INFO - Iter 4200/10000, Loss: 0.0465, Gradient Norm: 1.6722, Av
g Adaptive LR: 0.21796820923788462
2025-03-18 21:34:39,392 - INFO - Iter 4300/10000, Loss: 0.0465, Gradient Norm: 1.6752, Av
g Adaptive LR: 0.21542520573852025
2025-03-18 21:34:42,757 - INFO - Iter 4400/10000, Loss: 0.0461, Gradient Norm: 1.6374, Av
g Adaptive LR: 0.2129679892850653
2025-03-18 21:34:46,091 - INFO - Iter 4500/10000, Loss: 0.0459, Gradient Norm: 1.6090, Av
g Adaptive LR: 0.21059383798786804
2025-03-18 21:34:49,442 - INFO - Iter 4600/10000, Loss: 0.0459, Gradient Norm: 1.6110, Av
g Adaptive LR: 0.20829727830070627
2025-03-18 21:34:52,783 - INFO - Iter 4700/10000, Loss: 0.0457, Gradient Norm: 1.5900, Av
g Adaptive LR: 0.20607421161521577
2025-03-18 21:34:56,082 - INFO - Iter 4800/10000, Loss: 0.0454, Gradient Norm: 1.5527, Av
g Adaptive LR: 0.20392121345025044
2025-03-18 21:34:59,357 - INFO - Iter 4900/10000, Loss: 0.0454, Gradient Norm: 1.5490, Av
g Adaptive LR: 0.2018342061289959
2025-03-18 21:35:02,694 - INFO - Iter 5000/10000, Loss: 0.0453, Gradient Norm: 1.5425, Av
g Adaptive LR: 0.19981038203533483
2025-03-18 21:35:05,921 - INFO - Iter 5100/10000, Loss: 0.0450, Gradient Norm: 1.5040, Av
g Adaptive LR: 0.19784547425270002
2025-03-18 21:35:09,175 - INFO - Iter 5200/10000, Loss: 0.0449, Gradient Norm: 1.4912, Av
g Adaptive LR: 0.19593783832206368
2025-03-18 21:35:12,372 - INFO - Iter 5300/10000, Loss: 0.0449, Gradient Norm: 1.4890, Av
g Adaptive LR: 0.19408426482384733
2025-03-18 21:35:15,645 - INFO - Iter 5400/10000, Loss: 0.0447, Gradient Norm: 1.4636, Av
g Adaptive LR: 0.19228256753784848
2025-03-18 21:35:18,938 - INFO - Iter 5500/10000, Loss: 0.0445, Gradient Norm: 1.4434, Av

g Adaptive LR: 0.19052980594452666
2025-03-18 21:35:22,155 - INFO - Iter 5600/10000, Loss: 0.0444, Gradient Norm: 1.4367, Av
g Adaptive LR: 0.18882503043912416
2025-03-18 21:35:25,481 - INFO - Iter 5700/10000, Loss: 0.0444, Gradient Norm: 1.4276, Av
g Adaptive LR: 0.18716436337953965
2025-03-18 21:35:28,761 - INFO - Iter 5800/10000, Loss: 0.0442, Gradient Norm: 1.4025, Av
g Adaptive LR: 0.18554718978001475
2025-03-18 21:35:32,037 - INFO - Iter 5900/10000, Loss: 0.0441, Gradient Norm: 1.3931, Av
g Adaptive LR: 0.1839707538697598
2025-03-18 21:35:35,321 - INFO - Iter 6000/10000, Loss: 0.0441, Gradient Norm: 1.3878, Av
g Adaptive LR: 0.18243476864433872
2025-03-18 21:35:38,594 - INFO - Iter 6100/10000, Loss: 0.0439, Gradient Norm: 1.3696, Av
g Adaptive LR: 0.18093584665879728
2025-03-18 21:35:41,876 - INFO - Iter 6200/10000, Loss: 0.0438, Gradient Norm: 1.3551, Av
g Adaptive LR: 0.17947355347817873
2025-03-18 21:35:45,220 - INFO - Iter 6300/10000, Loss: 0.0438, Gradient Norm: 1.3478, Av
g Adaptive LR: 0.17804612423105168
2025-03-18 21:35:48,639 - INFO - Iter 6400/10000, Loss: 0.0437, Gradient Norm: 1.3430, Av
g Adaptive LR: 0.1766524034147619
2025-03-18 21:35:51,953 - INFO - Iter 6500/10000, Loss: 0.0436, Gradient Norm: 1.3226, Av
g Adaptive LR: 0.17529091647971973
2025-03-18 21:35:55,239 - INFO - Iter 6600/10000, Loss: 0.0435, Gradient Norm: 1.3150, Av
g Adaptive LR: 0.17396017488703852
2025-03-18 21:35:58,547 - INFO - Iter 6700/10000, Loss: 0.0435, Gradient Norm: 1.3079, Av
g Adaptive LR: 0.17265966822877774
2025-03-18 21:36:01,901 - INFO - Iter 6800/10000, Loss: 0.0434, Gradient Norm: 1.3028, Av
g Adaptive LR: 0.17138759271534149
2025-03-18 21:36:05,203 - INFO - Iter 6900/10000, Loss: 0.0433, Gradient Norm: 1.2837, Av
g Adaptive LR: 0.17014353911103977
2025-03-18 21:36:08,416 - INFO - Iter 7000/10000, Loss: 0.0433, Gradient Norm: 1.2790, Av
g Adaptive LR: 0.1689258413032433
2025-03-18 21:36:11,623 - INFO - Iter 7100/10000, Loss: 0.0432, Gradient Norm: 1.2720, Av
g Adaptive LR: 0.16773434925071556
2025-03-18 21:36:14,889 - INFO - Iter 7200/10000, Loss: 0.0432, Gradient Norm: 1.2672, Av
g Adaptive LR: 0.16656741464314984
2025-03-18 21:36:18,314 - INFO - Iter 7300/10000, Loss: 0.0430, Gradient Norm: 1.2501, Av
g Adaptive LR: 0.16542463890351247
2025-03-18 21:36:21,606 - INFO - Iter 7400/10000, Loss: 0.0430, Gradient Norm: 1.2457, Av
g Adaptive LR: 0.1643049756591076
2025-03-18 21:36:24,925 - INFO - Iter 7500/10000, Loss: 0.0430, Gradient Norm: 1.2385, Av
g Adaptive LR: 0.1632079140985682
2025-03-18 21:36:28,142 - INFO - Iter 7600/10000, Loss: 0.0429, Gradient Norm: 1.2345, Av
g Adaptive LR: 0.16213250277294099
2025-03-18 21:36:31,458 - INFO - Iter 7700/10000, Loss: 0.0428, Gradient Norm: 1.2211, Av
g Adaptive LR: 0.16107787099036086
2025-03-18 21:36:34,813 - INFO - Iter 7800/10000, Loss: 0.0428, Gradient Norm: 1.2145, Av
g Adaptive LR: 0.16004389032051447
2025-03-18 21:36:38,092 - INFO - Iter 7900/10000, Loss: 0.0427, Gradient Norm: 1.2070, Av
g Adaptive LR: 0.15902934769847135
2025-03-18 21:36:41,339 - INFO - Iter 8000/10000, Loss: 0.0427, Gradient Norm: 1.2034, Av
g Adaptive LR: 0.1580340102334851
2025-03-18 21:36:44,674 - INFO - Iter 8100/10000, Loss: 0.0427, Gradient Norm: 1.1960, Av
g Adaptive LR: 0.15705695610895032
2025-03-18 21:36:47,957 - INFO - Iter 8200/10000, Loss: 0.0426, Gradient Norm: 1.1851, Av
g Adaptive LR: 0.1560981171657824
2025-03-18 21:36:51,260 - INFO - Iter 8300/10000, Loss: 0.0425, Gradient Norm: 1.1770, Av
g Adaptive LR: 0.15515632620509023
2025-03-18 21:36:54,594 - INFO - Iter 8400/10000, Loss: 0.0425, Gradient Norm: 1.1734, Av
g Adaptive LR: 0.15423144664948793
2025-03-18 21:36:57,872 - INFO - Iter 8500/10000, Loss: 0.0425, Gradient Norm: 1.1704, Av
g Adaptive LR: 0.15332317903796416
2025-03-18 21:37:01,142 - INFO - Iter 8600/10000, Loss: 0.0424, Gradient Norm: 1.1590, Av
g Adaptive LR: 0.15243043136164477
2025-03-18 21:37:04,410 - INFO - Iter 8700/10000, Loss: 0.0424, Gradient Norm: 1.1487, Av
g Adaptive LR: 0.15155330797529604
2025-03-18 21:37:07,740 - INFO - Iter 8800/10000, Loss: 0.0423, Gradient Norm: 1.1443, Av
g Adaptive LR: 0.1506909962992794
2025-03-18 21:37:10,952 - INFO - Iter 8900/10000, Loss: 0.0423, Gradient Norm: 1.1428, Av
g Adaptive LR: 0.14984351124900186
2025-03-18 21:37:14,230 - INFO - Iter 9000/10000, Loss: 0.0423, Gradient Norm: 1.1380, Av
g Adaptive LR: 0.14900989862997024
2025-03-18 21:37:17,491 - INFO - Iter 9100/10000, Loss: 0.0422, Gradient Norm: 1.1235, Av

```
g Adaptive LR: 0.1481901965458283
2025-03-18 21:37:20,754 - INFO - Iter 9200/10000, Loss: 0.0421, Gradient Norm: 1.1157, Av
g Adaptive LR: 0.1473839649202619
2025-03-18 21:37:24,001 - INFO - Iter 9300/10000, Loss: 0.0421, Gradient Norm: 1.1150, Av
g Adaptive LR: 0.14659045143428331
2025-03-18 21:37:27,226 - INFO - Iter 9400/10000, Loss: 0.0421, Gradient Norm: 1.1151, Av
g Adaptive LR: 0.14581009826726885
2025-03-18 21:37:30,552 - INFO - Iter 9500/10000, Loss: 0.0421, Gradient Norm: 1.1039, Av
g Adaptive LR: 0.14504150676819993
2025-03-18 21:37:33,828 - INFO - Iter 9600/10000, Loss: 0.0420, Gradient Norm: 1.0895, Av
g Adaptive LR: 0.14428549327819923
2025-03-18 21:37:37,140 - INFO - Iter 9700/10000, Loss: 0.0420, Gradient Norm: 1.0875, Av
g Adaptive LR: 0.1435407660070575
2025-03-18 21:37:40,407 - INFO - Iter 9800/10000, Loss: 0.0420, Gradient Norm: 1.0886, Av
g Adaptive LR: 0.14280771257122823
2025-03-18 21:37:43,709 - INFO - Iter 9900/10000, Loss: 0.0420, Gradient Norm: 1.0871, Av
g Adaptive LR: 0.14208573850875456
2025-03-18 21:37:46,959 - INFO - Iter 10000/10000, Loss: 0.0419, Gradient Norm: 1.0712, A
vg Adaptive LR: 0.1413744125522334
2025-03-18 21:37:46,959 - INFO - LinearRegressionClassifier training completed in 330.34
seconds.
Train Regressions: 100%|██████████| 3/3 [11:58<00:00, 239.65s/it]
2025-03-18 21:37:46,961 - INFO - Training complete for Softmax and Linear.
2025-03-18 21:37:46,961 - INFO - === TRAINING PERCEPTRON MODELS (Clean & Pocket) ===
Train Clean & Pocket:   0%|          | 0/4 [00:00<?, ?it/s]2025-03-18 21:37:46,963 - INFO
- --- Clean PLA, max_iter=20 ---
2025-03-18 21:37:46,964 - INFO - Training for digit 0...
2025-03-18 21:37:47,972 - INFO - Training for digit 1...
2025-03-18 21:37:49,034 - INFO - Training for digit 2...
2025-03-18 21:37:50,254 - INFO - Training for digit 3...
2025-03-18 21:37:51,522 - INFO - Training for digit 4...
2025-03-18 21:37:52,763 - INFO - Training for digit 5...
2025-03-18 21:37:54,106 - INFO - Training for digit 6...
2025-03-18 21:37:55,176 - INFO - Training for digit 7...
2025-03-18 21:37:56,322 - INFO - Training for digit 8...
2025-03-18 21:37:57,994 - INFO - Training for digit 9...
2025-03-18 21:37:59,592 - INFO - --- Pocket PLA, max_iter=20 ---
2025-03-18 21:37:59,592 - INFO - Training for digit 0...
2025-03-18 21:38:00,650 - INFO - Training for digit 1...
2025-03-18 21:38:01,730 - INFO - Training for digit 2...
2025-03-18 21:38:02,961 - INFO - Training for digit 3...
2025-03-18 21:38:04,221 - INFO - Training for digit 4...
2025-03-18 21:38:05,468 - INFO - Training for digit 5...
2025-03-18 21:38:06,773 - INFO - Training for digit 6...
2025-03-18 21:38:07,872 - INFO - Training for digit 7...
2025-03-18 21:38:09,031 - INFO - Training for digit 8...
2025-03-18 21:38:10,655 - INFO - Training for digit 9...
Train Clean & Pocket:  25%|██        | 1/4 [00:25<01:15, 25.32s/it]2025-03-18 21:38:12,28
0 - INFO - --- Clean PLA, max_iter=50 ---
2025-03-18 21:38:12,280 - INFO - Training for digit 0...
2025-03-18 21:38:14,039 - INFO - Training for digit 1...
2025-03-18 21:38:15,799 - INFO - Training for digit 2...
2025-03-18 21:38:18,045 - INFO - Training for digit 3...
2025-03-18 21:38:20,397 - INFO - Training for digit 4...
2025-03-18 21:38:22,559 - INFO - Training for digit 5...
2025-03-18 21:38:25,615 - INFO - Training for digit 6...
2025-03-18 21:38:27,543 - INFO - Training for digit 7...
2025-03-18 21:38:29,568 - INFO - Training for digit 8...
2025-03-18 21:38:33,354 - INFO - Training for digit 9...
2025-03-18 21:38:36,992 - INFO - --- Pocket PLA, max_iter=50 ---
2025-03-18 21:38:36,993 - INFO - Training for digit 0...
2025-03-18 21:38:38,863 - INFO - Training for digit 1...
2025-03-18 21:38:40,743 - INFO - Training for digit 2...
2025-03-18 21:38:43,051 - INFO - Training for digit 3...
2025-03-18 21:38:45,422 - INFO - Training for digit 4...
2025-03-18 21:38:47,629 - INFO - Training for digit 5...
2025-03-18 21:38:50,654 - INFO - Training for digit 6...
2025-03-18 21:38:52,562 - INFO - Training for digit 7...
2025-03-18 21:38:54,551 - INFO - Training for digit 8...
2025-03-18 21:38:58,185 - INFO - Training for digit 9...
Train Clean & Pocket:  50%|█████     | 2/4 [01:14<01:18, 39.48s/it]2025-03-18 21:39:01,67
4 - INFO - --- Clean PLA, max_iter=100 ---
```

```
2025-03-18 21:39:01,674 - INFO - Training for digit 0...
2025-03-18 21:39:04,555 - INFO - Training for digit 1...
2025-03-18 21:39:07,416 - INFO - Training for digit 2...
2025-03-18 21:39:11,218 - INFO - Training for digit 3...
2025-03-18 21:39:15,634 - INFO - Training for digit 4...
2025-03-18 21:39:19,287 - INFO - Training for digit 5...
2025-03-18 21:39:24,975 - INFO - Training for digit 6...
2025-03-18 21:39:28,103 - INFO - Training for digit 7...
2025-03-18 21:39:31,379 - INFO - Training for digit 8...
2025-03-18 21:39:38,496 - INFO - Training for digit 9...
2025-03-18 21:39:44,509 - INFO - --- Pocket PLA, max_iter=100 ---
2025-03-18 21:39:44,510 - INFO - Training for digit 0...
2025-03-18 21:39:47,377 - INFO - Training for digit 1...
2025-03-18 21:39:50,199 - INFO - Training for digit 2...
2025-03-18 21:39:53,919 - INFO - Training for digit 3...
2025-03-18 21:39:58,213 - INFO - Training for digit 4...
2025-03-18 21:40:01,847 - INFO - Training for digit 5...
2025-03-18 21:40:07,546 - INFO - Training for digit 6...
2025-03-18 21:40:10,678 - INFO - Training for digit 7...
2025-03-18 21:40:13,978 - INFO - Training for digit 8...
2025-03-18 21:40:21,091 - INFO - Training for digit 9...
Train Clean & Pocket:  75%|███████▌  | 3/4 [02:40<01:00, 60.47s/it]2025-03-18 21:40:27,12
6 - INFO - --- Clean PLA, max_iter=1000 ---
2025-03-18 21:40:27,126 - INFO - Training for digit 0...
2025-03-18 21:40:47,024 - INFO - Training for digit 1...
2025-03-18 21:41:07,128 - INFO - Training for digit 2...
2025-03-18 21:41:39,843 - INFO - Training for digit 3...
2025-03-18 21:42:18,041 - INFO - Training for digit 4...
2025-03-18 21:42:46,958 - INFO - Training for digit 5...
2025-03-18 21:43:30,948 - INFO - Training for digit 6...
2025-03-18 21:43:55,779 - INFO - Training for digit 7...
2025-03-18 21:44:22,543 - INFO - Training for digit 8...
2025-03-18 21:45:20,724 - INFO - Training for digit 9...
2025-03-18 21:46:10,398 - INFO - --- Pocket PLA, max_iter=1000 ---
2025-03-18 21:46:10,398 - INFO - Training for digit 0...
2025-03-18 21:46:30,402 - INFO - Training for digit 1...
2025-03-18 21:46:50,447 - INFO - Training for digit 2...
2025-03-18 21:47:22,917 - INFO - Training for digit 3...
2025-03-18 21:48:01,110 - INFO - Training for digit 4...
2025-03-18 21:48:30,042 - INFO - Training for digit 5...
2025-03-18 21:49:14,038 - INFO - Training for digit 6...
2025-03-18 21:49:38,992 - INFO - Training for digit 7...
2025-03-18 21:50:05,786 - INFO - Training for digit 8...
2025-03-18 21:51:03,991 - INFO - Training for digit 9...
Train Clean & Pocket: 100%|██████████| 4/4 [14:06<00:00, 211.75s/it]
2025-03-18 21:51:53,952 - INFO - Training complete for Clean PLA and Pocket PLA.
2025-03-18 21:51:53,953 - INFO - === ALL TRAINING COMPLETE ===
```

# Evaluate

In [7]:

```
################################################
# EVALUATION CELL (with pandas DataFrame)
################################################


# 1) Evaluate Perceptrons: Clean & Pocket
accuracies_clean, accuracies_pocket = [], []
runtimes_clean,   runtimes_pocket   = [], []
sensitivities_clean, sensitivities_pocket = [], []
selectivities_clean, selectivities_pocket = [], []

conf_clean, conf_pocket = [], []
meta_clean, meta_pocket = [], []

for max_iter in tqdm(perceptron_max_iter_values, desc="Evaluate Clean & Pocket"):
    # === Evaluate Clean PLA ===
    c_model = trained_models_clean[max_iter]
    cm_c, acc_c, s_c, sp_c, rt_c, ex_c = evaluate_model(
```

```
        c_model, X_test, y_test, classes=range(10), model_name="Clean PLA"
    )
    accuracies_clean.append(acc_c)
    runtimes_clean.append(rt_c)
    sensitivities_clean.append(np.mean(s_c))
    selectivities_clean.append(np.mean(sp_c))
    conf_clean.append(cm_c)

    cdict = {
        "max_iter": max_iter,
        "accuracy": acc_c,
        "runtime": rt_c,
        "avg_sensitivity": np.mean(s_c),
        "avg_selectivity": np.mean(sp_c),
        "method": "Clean PLA"
    }
    cdict.update(ex_c)
    meta_clean.append(cdict)

    # === Evaluate Pocket PLA ===
    p_model = trained_models_pocket[max_iter]
    cm_p, acc_p, s_p, sp_p, rt_p, ex_p = evaluate_model(
        p_model, X_test, y_test, classes=range(10), model_name="Pocket PLA"
    )
    accuracies_pocket.append(acc_p)
    runtimes_pocket.append(rt_p)
    sensitivities_pocket.append(np.mean(s_p))
    selectivities_pocket.append(np.mean(sp_p))
    conf_pocket.append(cm_p)

    pdict = {
        "max_iter": max_iter,
        "accuracy": acc_p,
        "runtime": rt_p,
        "avg_sensitivity": np.mean(s_p),
        "avg_selectivity": np.mean(sp_p),
        "method": "Pocket PLA"
    }
    pdict.update(ex_p)
    meta_pocket.append(pdict)

# Aggregated iteration-level training curves for Perceptrons
clean_train_curve = aggregate_iteration_losses(
    [trained_models_clean[m] for m in perceptron_max_iter_values]
)
pocket_train_curve = aggregate_iteration_losses(
    [trained_models_pocket[m] for m in perceptron_max_iter_values]
)

# 2) Evaluate Regression Models: Softmax & Linear
accuracies_softmax = []
runtimes_softmax   = []
sensitivities_soft = []
selectivities_soft = []
conf_soft          = []
meta_soft          = []

accuracies_linear = []
runtimes_linear   = []
sensitivities_lin = []
selectivities_lin = []
conf_linear       = []
meta_linear       = []

for cfg in tqdm(regression_run_configs, desc="Evaluate Regressions"):
    lr_val = cfg["learning_rate"]
    max_iter_val = cfg["max_iter"]
    label = cfg["label"]

    # === Evaluate Softmax ===
    s_model = trained_models_softmax[(lr_val, max_iter_val)]
    cm_s, a_s, se_s, sp_s, r_s, ex_s = evaluate_model(
```

```
        s_model, X_test, y_test, classes=range(10),
        model_name=f"Softmax ({label})"
    )
    accuracies_softmax.append(a_s)
    runtimes_softmax.append(r_s)
    sensitivities_soft.append(np.mean(se_s))
    selectivities_soft.append(np.mean(sp_s))
    conf_soft.append(cm_s)

    ms = {
        "label": label,
        "learning_rate": lr_val,
        "max_iter": max_iter_val,
        "accuracy": a_s,
        "runtime": r_s,
        "avg_sensitivity": np.mean(se_s),
        "avg_selectivity": np.mean(sp_s),
        "method": "Softmax"
    }
    ms.update(ex_s)
    meta_soft.append(ms)

    # === Evaluate Linear ===
    lin_model = trained_models_linear[(lr_val, max_iter_val)]
    cm_l, a_l, se_l, sp_l, r_l, ex_l = evaluate_model(
        lin_model, X_test, y_test, classes=range(10),
        model_name=f"Linear ({label})"
    )
    accuracies_linear.append(a_l)
    runtimes_linear.append(r_l)
    sensitivities_lin.append(np.mean(se_l))
    selectivities_lin.append(np.mean(sp_l))
    conf_linear.append(cm_l)

    ml = {
        "label": label,
        "learning_rate": lr_val,
        "max_iter": max_iter_val,
        "accuracy": a_l,
        "runtime": r_l,
        "avg_sensitivity": np.mean(se_l),
        "avg_selectivity": np.mean(sp_l),
        "method": "Linear Regression"
    }
    ml.update(ex_l)
    meta_linear.append(ml)


logger.info("Evaluation complete for Perceptrons & Regressions.")


# # 1) Build the DataFrame of all model results
# all_rows = []

# # A) Clean PLA
# for i, max_iter in tqdm(
#     enumerate(perceptron_max_iter_values),
#     desc="Collecting Clean PLA",
#     total=len(perceptron_max_iter_values)
# ):
#     all_rows.append({
#         'model': 'Clean PLA',
#         'max_iter': max_iter,
#         'runtime': runtimes_clean[i],
#         'accuracy': accuracies_clean[i],
#         'sensitivity': sensitivities_clean[i],
#         'selectivity': selectivities_clean[i]
#     })

# # B) Pocket PLA
# for i, max_iter in tqdm(
#     enumerate(perceptron_max_iter_values),
```

```
#     desc="Collecting Pocket PLA",
#     total=len(perceptron_max_iter_values)
# ):
#     all_rows.append({
#         'model': 'Pocket PLA',
#         'max_iter': max_iter,
#         'runtime': runtimes_pocket[i],
#         'accuracy': accuracies_pocket[i],
#         'sensitivity': sensitivities_pocket[i],
#         'selectivity': selectivities_pocket[i]
#     })

# # C) Softmax
# for i, row_meta in tqdm(
#     enumerate(meta_soft),
#     desc="Collecting Softmax",
#     total=len(meta_soft)
# ):
#     all_rows.append({
#         'model': 'Softmax',
#         'max_iter': row_meta['max_iter'],
#         'runtime': runtimes_softmax[i],
#         'accuracy': accuracies_softmax[i],
#         'sensitivity': sensitivities_soft[i],
#         'selectivity': selectivities_soft[i]
#     })

# # D) Linear
# for i, row_meta in tqdm(
#     enumerate(meta_linear),
#     desc="Collecting Linear",
#     total=len(meta_linear)
# ):
#     all_rows.append({
#         'model': 'Linear',
#         'max_iter': row_meta['max_iter'],
#         'runtime': runtimes_linear[i],
#         'accuracy': accuracies_linear[i],
#         'sensitivity': sensitivities_lin[i],
#         'selectivity': selectivities_lin[i]
#     })
```

```
Evaluate Clean & Pocket:   0%|          | 0/4 [00:00<?, ?it/s]2025-03-18 21:51:53,971 - I
NFO - Built-in Confusion Matrix:
[[ 941    0    4    4    0   19   11    1    0    0]
 [   0 1066   31   17    1   16    4    0    0    0]
 [  20    6  899   38   19    5   26   17    2    0]
 [   8    1   31  901    1   40    7   19    0    2]
 [   4    7    8    3  908   15   23    5    2    7]
 [  33    6   13   73   19  716   21    8    0    3]
 [  22    5   14    2   10   26  879    0    0    0]
 [   7   23   49    6   10    1    4  923    0    5]
 [  12   17   93  181   14  471   25   32  121    8]
 [  18    9   27   22  293   80    5  174    0  381]]
2025-03-18 21:51:53,972 - INFO - Overall Accuracy: 77.35%
2025-03-18 21:51:53,971 - INFO - Built-in Confusion Matrix:
[[ 941    0    4    4    0   19   11    1    0    0]
 [   0 1066   31   17    1   16    4    0    0    0]
 [  20    6  899   38   19    5   26   17    2    0]
 [   8    1   31  901    1   40    7   19    0    2]
 [   4    7    8    3  908   15   23    5    2    7]
 [  33    6   13   73   19  716   21    8    0    3]
 [  22    5   14    2   10   26  879    0    0    0]
 [   7   23   49    6   10    1    4  923    0    5]
 [  12   17   93  181   14  471   25   32  121    8]
 [  18    9   27   22  293   80    5  174    0  381]]
2025-03-18 21:51:53,972 - INFO - Overall Accuracy: 77.35%
2025-03-18 21:51:53,974 - INFO - Class '0': TPR=0.96, TNR=0.99
2025-03-18 21:51:53,974 - INFO - Class '1': TPR=0.94, TNR=0.99
2025-03-18 21:51:53,974 - INFO - Class '2': TPR=0.87, TNR=0.97
2025-03-18 21:51:53,975 - INFO - Class '3': TPR=0.89, TNR=0.96
2025-03-18 21:51:53,976 - INFO - Class '4': TPR=0.92, TNR=0.96
```

```
2025-03-18 21:51:53,976 - INFO - Class '5': TPR=0.80, TNR=0.93
2025-03-18 21:51:53,977 - INFO - Class '6': TPR=0.92, TNR=0.99
2025-03-18 21:51:53,977 - INFO - Class '7': TPR=0.90, TNR=0.97
2025-03-18 21:51:53,978 - INFO - Class '8': TPR=0.12, TNR=1.00
2025-03-18 21:51:53,978 - INFO - Class '9': TPR=0.38, TNR=1.00
Evaluating class metrics: 100%|██████████| 10/10 [00:00<00:00, 2132.12it/s]
2025-03-18 21:51:53,983 - INFO - Built-in Confusion Matrix:
[[ 940    0    4    4    0   13   11    1    7    0]
 [   0 1062   19   14    1   14    4    0   21    0]
 [  20    6  880   33   19    2   26   17   29    0]
 [   8    1   28  895    1   38    7   17   13    2]
 [   4    6    7    4  903   15   23    5   12    3]
 [  33    6    9   73   17  696   19    8   29    2]
 [  22    4   14    2   10   25  876    0    5    0]
 [   7   23   46    7   10    0    4  921    5    5]
 [   9    5   24   72    5   68   16   17  758    0]
 [  19    8   27   19  237   75    4  138   36  446]]
2025-03-18 21:51:53,983 - INFO - Overall Accuracy: 83.77%
2025-03-18 21:51:53,986 - INFO - Class '0': TPR=0.96, TNR=0.99
2025-03-18 21:51:53,986 - INFO - Class '1': TPR=0.94, TNR=0.99
2025-03-18 21:51:53,987 - INFO - Class '2': TPR=0.85, TNR=0.98
2025-03-18 21:51:53,987 - INFO - Class '3': TPR=0.89, TNR=0.97
2025-03-18 21:51:53,988 - INFO - Class '4': TPR=0.92, TNR=0.97
2025-03-18 21:51:53,988 - INFO - Class '5': TPR=0.78, TNR=0.97
2025-03-18 21:51:53,988 - INFO - Class '6': TPR=0.91, TNR=0.99
2025-03-18 21:51:53,989 - INFO - Class '7': TPR=0.90, TNR=0.98
2025-03-18 21:51:53,989 - INFO - Class '8': TPR=0.78, TNR=0.98
2025-03-18 21:51:53,989 - INFO - Class '9': TPR=0.44, TNR=1.00
Evaluating class metrics: 100%|██████████| 10/10 [00:00<00:00, 2506.76it/s]
2025-03-18 21:51:53,994 - INFO - Built-in Confusion Matrix:
[[ 959    0    3    6    1    0    7    1    3    0]
 [   0 1077   18   23    1    0    4    2   10    0]
 [   9    7  912   26   16    0   20   20   19    3]
 [   6    0   20  955    1    4    4   13    6    1]
 [   1    0    8    7  933    0   14    2    4   13]
 [  33    6   15  169   42  526   27   18   45   11]
 [  19    3    9    7   11    3  905    0    1    0]
 [   6   13   44   13    9    0    2  936    0    5]
 [  18    7   39  137   18    8   18   23  704    2]
 [  13    8   23   47  174    5    0   98    3  638]]
2025-03-18 21:51:53,994 - INFO - Overall Accuracy: 85.45%
2025-03-18 21:51:53,997 - INFO - Class '0': TPR=0.98, TNR=0.99
2025-03-18 21:51:53,997 - INFO - Class '1': TPR=0.95, TNR=1.00
2025-03-18 21:51:53,997 - INFO - Class '2': TPR=0.88, TNR=0.98
2025-03-18 21:51:53,998 - INFO - Class '3': TPR=0.95, TNR=0.95
2025-03-18 21:51:53,998 - INFO - Class '4': TPR=0.95, TNR=0.97
2025-03-18 21:51:53,999 - INFO - Class '5': TPR=0.59, TNR=1.00
2025-03-18 21:51:53,999 - INFO - Class '6': TPR=0.94, TNR=0.99
2025-03-18 21:51:54,000 - INFO - Class '7': TPR=0.91, TNR=0.98
2025-03-18 21:51:54,000 - INFO - Class '8': TPR=0.72, TNR=0.99
2025-03-18 21:51:54,001 - INFO - Class '9': TPR=0.63, TNR=1.00
Evaluating class metrics: 100%|██████████| 10/10 [00:00<00:00, 2116.20it/s]
2025-03-18 21:51:54,005 - INFO - Built-in Confusion Matrix:
[[ 956    0    3    3    1    0    6    1   10    0]
 [   0 1075   14   18    1    2    4    1   20    0]
 [   9    5  901   21   15    0   18   18   42    3]
 [   6    0   20  928    1   17    4   12   21    1]
 [   1    0    8    6  924    1   13    2   13   14]
 [  20    4    9   88   20  653   20   10   64    4]
 [  16    3    7    5   11   12  892    0   12    0]
 [   5   12   44   11    8    0    2  934    5    7]
 [  11    3   19   58    7   16    9   14  837    0]
 [  12    7   21   31  149   17    0   83   36  653]]
2025-03-18 21:51:54,006 - INFO - Overall Accuracy: 87.53%
2025-03-18 21:51:54,008 - INFO - Class '0': TPR=0.98, TNR=0.99
2025-03-18 21:51:54,009 - INFO - Class '1': TPR=0.95, TNR=1.00
2025-03-18 21:51:54,009 - INFO - Class '2': TPR=0.87, TNR=0.98
2025-03-18 21:51:54,010 - INFO - Class '3': TPR=0.92, TNR=0.97
2025-03-18 21:51:54,010 - INFO - Class '4': TPR=0.94, TNR=0.98
2025-03-18 21:51:54,011 - INFO - Class '5': TPR=0.73, TNR=0.99
2025-03-18 21:51:54,011 - INFO - Class '6': TPR=0.93, TNR=0.99
2025-03-18 21:51:54,011 - INFO - Class '7': TPR=0.91, TNR=0.98
```

```
2025-03-18 21:51:54,012 - INFO - Class '8': TPR=0.86, TNR=0.98
2025-03-18 21:51:54,013 - INFO - Class '9': TPR=0.65, TNR=1.00
Evaluating class metrics: 100%|██████████| 10/10 [00:00<00:00, 2124.13it/s]
2025-03-18 21:51:54,017 - INFO - Built-in Confusion Matrix:
[[ 964    0    3    2    1    1    6    2    1    0]
 [   0 1107   10    6    0    2    5    2    3    0]
 [  18   12  914    9   13    1   23   20   15    7]
 [  12    1   26  910    2   20    6   18    3   12]
 [   2    1    5    0  930    0   11    3    2   28]
 [  25    6   13   44   33  703   30   18    8   12]
 [  12    3    5    2   10    6  920    0    0    0]
 [   5    8   29    5    7    0    2  951    0   21]
 [  30   14  105   81   59   44   29   32  542   38]
 [  12    7   10   17   93    5    1   60    0  804]]
2025-03-18 21:51:54,017 - INFO - Overall Accuracy: 87.45%
2025-03-18 21:51:54,019 - INFO - Class '0': TPR=0.98, TNR=0.99
2025-03-18 21:51:54,020 - INFO - Class '1': TPR=0.98, TNR=0.99
2025-03-18 21:51:54,020 - INFO - Class '2': TPR=0.89, TNR=0.98
2025-03-18 21:51:54,021 - INFO - Class '3': TPR=0.90, TNR=0.98
2025-03-18 21:51:54,021 - INFO - Class '4': TPR=0.95, TNR=0.98
2025-03-18 21:51:54,022 - INFO - Class '5': TPR=0.79, TNR=0.99
2025-03-18 21:51:54,022 - INFO - Class '6': TPR=0.96, TNR=0.99
2025-03-18 21:51:54,023 - INFO - Class '7': TPR=0.93, TNR=0.98
2025-03-18 21:51:54,023 - INFO - Class '8': TPR=0.56, TNR=1.00
2025-03-18 21:51:54,023 - INFO - Class '9': TPR=0.80, TNR=0.99
Evaluating class metrics: 100%|██████████| 10/10 [00:00<00:00, 2162.57it/s]
2025-03-18 21:51:54,027 - INFO - Built-in Confusion Matrix:
[[ 963    0    3    3    1    0    5    2    3    0]
 [   0 1097    9    6    0    1    4    1   17    0]
 [   8    3  906   20   12    0   16   17   43    7]
 [   6    0   21  921    1   18    4   13   19    7]
 [   2    0    8    2  916    1    9    2   11   31]
 [  21    4   10   65   24  664   22   14   58   10]
 [  12    3    9    3   10    7  909    0    5    0]
 [   5    7   32    9    6    0    2  943    2   22]
 [  13    3   24   51   14   12   14   17  821    5]
 [  10    7   11   20   70   10    0   46   11  824]]
2025-03-18 21:51:54,028 - INFO - Overall Accuracy: 89.64%
2025-03-18 21:51:54,030 - INFO - Class '0': TPR=0.98, TNR=0.99
2025-03-18 21:51:54,030 - INFO - Class '1': TPR=0.97, TNR=1.00
2025-03-18 21:51:54,031 - INFO - Class '2': TPR=0.88, TNR=0.99
2025-03-18 21:51:54,031 - INFO - Class '3': TPR=0.91, TNR=0.98
2025-03-18 21:51:54,031 - INFO - Class '4': TPR=0.93, TNR=0.98
2025-03-18 21:51:54,032 - INFO - Class '5': TPR=0.74, TNR=0.99
2025-03-18 21:51:54,032 - INFO - Class '6': TPR=0.95, TNR=0.99
2025-03-18 21:51:54,033 - INFO - Class '7': TPR=0.92, TNR=0.99
2025-03-18 21:51:54,033 - INFO - Class '8': TPR=0.84, TNR=0.98
2025-03-18 21:51:54,034 - INFO - Class '9': TPR=0.82, TNR=0.99
Evaluating class metrics: 100%|██████████| 10/10 [00:00<00:00, 2186.47it/s]
2025-03-18 21:51:54,038 - INFO - Built-in Confusion Matrix:
[[ 949    0    7    3    0    5   12    3    1    0]
 [   0 1100   24    2    0    2    4    2    1    0]
 [   9    2  968   10   10    2   13    8    7    3]
 [   4    2   46  915    1   24    3   12    3    0]
 [   2    3   14    4  934    0    8    6    3    8]
 [  15    4   22   37   12  773   13    6    9    1]
 [  12    3   23    1    4   19  894    2    0    0]
 [   5    5   47    6    9    2    1  949    0    4]
 [  43   29  148   50   38   74    7   34  551    0]
 [  24   13   18   21  132   20    0  181    0  600]]
2025-03-18 21:51:54,039 - INFO - Overall Accuracy: 86.33%
2025-03-18 21:51:54,041 - INFO - Class '0': TPR=0.97, TNR=0.99
2025-03-18 21:51:54,042 - INFO - Class '1': TPR=0.97, TNR=0.99
2025-03-18 21:51:54,042 - INFO - Class '2': TPR=0.94, TNR=0.96
2025-03-18 21:51:54,043 - INFO - Class '3': TPR=0.91, TNR=0.99
2025-03-18 21:51:54,043 - INFO - Class '4': TPR=0.95, TNR=0.98
2025-03-18 21:51:54,044 - INFO - Class '5': TPR=0.87, TNR=0.98
2025-03-18 21:51:54,044 - INFO - Class '6': TPR=0.93, TNR=0.99
2025-03-18 21:51:54,044 - INFO - Class '7': TPR=0.92, TNR=0.97
2025-03-18 21:51:54,045 - INFO - Class '8': TPR=0.57, TNR=1.00
2025-03-18 21:51:54,046 - INFO - Class '9': TPR=0.59, TNR=1.00
Evaluating class metrics: 100%|██████████| 10/10 [00:00<00:00, 2121.76it/s]
```

```
2025-03-18 21:51:54,050 - INFO - Built-in Confusion Matrix:
[[ 961    0    0    2    0    4    7    3    3    0]
 [   0 1110    3    2    0    3    5    2   10    0]
 [  12    5  926   18    7    6   17   14   22    5]
 [   6    2   21  914    2   29    6   11   12    7]
 [   3    2    7    4  908    2    9    6    5   36]
 [  14    3    5   33    9  776   19    4   23    6]
 [  13    3    6    1    9   18  906    1    1    0]
 [   5    8   24    8    7    3    1  943    1   28]
 [  22   21   14   35   30   49   17   26  754    6]
 [  19   10    1   14   48   14    0   54    2  847]]
2025-03-18 21:51:54,050 - INFO - Overall Accuracy: 90.45%
2025-03-18 21:51:54,053 - INFO - Class '0': TPR=0.98, TNR=0.99
2025-03-18 21:51:54,053 - INFO - Class '1': TPR=0.98, TNR=0.99
2025-03-18 21:51:54,054 - INFO - Class '2': TPR=0.90, TNR=0.99
2025-03-18 21:51:54,054 - INFO - Class '3': TPR=0.90, TNR=0.99
2025-03-18 21:51:54,055 - INFO - Class '4': TPR=0.92, TNR=0.99
2025-03-18 21:51:54,055 - INFO - Class '5': TPR=0.87, TNR=0.99
2025-03-18 21:51:54,055 - INFO - Class '6': TPR=0.95, TNR=0.99
2025-03-18 21:51:54,056 - INFO - Class '7': TPR=0.92, TNR=0.99
2025-03-18 21:51:54,056 - INFO - Class '8': TPR=0.77, TNR=0.99
2025-03-18 21:51:54,057 - INFO - Class '9': TPR=0.84, TNR=0.99
Evaluating class metrics: 100%|██████████| 10/10 [00:00<00:00, 2220.38it/s]
Evaluate Clean & Pocket: 100%|██████████| 4/4 [00:00<00:00, 44.08it/s]
Aggregating train losses across Perceptron models: 100%|██████████| 4/4 [00:00<00:00, 133
8.32it/s]
Aggregating train losses across Perceptron models: 100%|██████████| 4/4 [00:00<00:00, 126
8.89it/s]
Evaluate Regressions:   0%|          | 0/3 [00:00<?, ?it/s]2025-03-18 21:51:54,076 - INFO
- Built-in Confusion Matrix:
[[ 956    0    0    2    0    3    9    3    7    0]
 [   0 1112    2    3    0    2    4    1   11    0]
 [  11   10  907   22   12    1   12   12   42    3]
 [   3    1   22  907    0   30    2   11   24   10]
 [   2    1    8    0  915    1   12    2   10   31]
 [   8    2    3   33   12  759   20   10   35   10]
 [  12    3    9    1    8    8  913    2    2    0]
 [   1   10   23    7    7    0    0  941    4   35]
 [   6    6    8   29    9   22   11   13  860   10]
 [  11    6    1   13   40    7    0   22    5  904]]
2025-03-18 21:51:54,077 - INFO - Overall Accuracy: 91.74%
2025-03-18 21:51:54,079 - INFO - Class '0': TPR=0.98, TNR=0.99
2025-03-18 21:51:54,080 - INFO - Class '1': TPR=0.98, TNR=1.00
2025-03-18 21:51:54,080 - INFO - Class '2': TPR=0.88, TNR=0.99
2025-03-18 21:51:54,081 - INFO - Class '3': TPR=0.90, TNR=0.99
2025-03-18 21:51:54,081 - INFO - Class '4': TPR=0.93, TNR=0.99
2025-03-18 21:51:54,082 - INFO - Class '5': TPR=0.85, TNR=0.99
2025-03-18 21:51:54,082 - INFO - Class '6': TPR=0.95, TNR=0.99
2025-03-18 21:51:54,082 - INFO - Class '7': TPR=0.92, TNR=0.99
2025-03-18 21:51:54,083 - INFO - Class '8': TPR=0.88, TNR=0.98
2025-03-18 21:51:54,083 - INFO - Class '9': TPR=0.90, TNR=0.99
Evaluating class metrics: 100%|██████████| 10/10 [00:00<00:00, 2348.96it/s]
2025-03-18 21:51:54,088 - INFO - Built-in Confusion Matrix:
[[ 941    0    0    3    1   10   24    0    0    1]
 [   0 1011    0   34   39   25   19    0    2    5]
 [  92    2  132  298   76   29  270    1    4  128]
 [  10    0    0  911    7   23   16    1    0   42]
 [   1    0    0    1  872    2   21    0    1   84]
 [  21    1    0  100   32  671   36    0    2   29]
 [  15    2    0    1   16   17  907    0    0    0]
 [  32   10    0   57   72    9   13  392    0  443]
 [  42    3    0  175   64  314   95    0   71  210]
 [  21    1    0   17   48    7    3    0    0  912]]
2025-03-18 21:51:54,089 - INFO - Overall Accuracy: 68.20%
2025-03-18 21:51:54,091 - INFO - Class '0': TPR=0.96, TNR=0.97
2025-03-18 21:51:54,091 - INFO - Class '1': TPR=0.89, TNR=1.00
2025-03-18 21:51:54,092 - INFO - Class '2': TPR=0.13, TNR=1.00
2025-03-18 21:51:54,092 - INFO - Class '3': TPR=0.90, TNR=0.92
2025-03-18 21:51:54,093 - INFO - Class '4': TPR=0.89, TNR=0.96
2025-03-18 21:51:54,093 - INFO - Class '5': TPR=0.75, TNR=0.95
2025-03-18 21:51:54,094 - INFO - Class '6': TPR=0.95, TNR=0.95
2025-03-18 21:51:54,094 - INFO - Class '7': TPR=0.38, TNR=1.00
```

```
2025-03-18 21:51:54,095 - INFO - Class '8': TPR=0.07, TNR=1.00
2025-03-18 21:51:54,095 - INFO - Class '9': TPR=0.90, TNR=0.90
Evaluating class metrics: 100%|██████████| 10/10 [00:00<00:00, 1989.33it/s]
2025-03-18 21:51:54,102 - INFO - Built-in Confusion Matrix:
[[ 959    0    0    2    0    6    8    2    2    1]
 [   0 1114    3    3    0    1    3    2    9    0]
 [   7    8  925   17   10    3   11    9   39    3]
 [   3    0   17  926    1   22    3    9   22    7]
 [   1    1    4    2  916    0   11    5   10   32]
 [   7    2    3   34    8  777   14   10   31    6]
 [  12    3    8    1    7   15  908    2    2    0]
 [   1    6   24    6    7    1    0  952    2   29]
 [   6    8    6   20    9   22    7   14  877    5]
 [  10    7    1   10   24    7    0   25    7  918]]
2025-03-18 21:51:54,103 - INFO - Overall Accuracy: 92.72%
2025-03-18 21:51:54,105 - INFO - Class '0': TPR=0.98, TNR=0.99
2025-03-18 21:51:54,106 - INFO - Class '1': TPR=0.98, TNR=1.00
2025-03-18 21:51:54,106 - INFO - Class '2': TPR=0.90, TNR=0.99
2025-03-18 21:51:54,107 - INFO - Class '3': TPR=0.92, TNR=0.99
2025-03-18 21:51:54,107 - INFO - Class '4': TPR=0.93, TNR=0.99
2025-03-18 21:51:54,107 - INFO - Class '5': TPR=0.87, TNR=0.99
2025-03-18 21:51:54,108 - INFO - Class '6': TPR=0.95, TNR=0.99
2025-03-18 21:51:54,108 - INFO - Class '7': TPR=0.93, TNR=0.99
2025-03-18 21:51:54,108 - INFO - Class '8': TPR=0.90, TNR=0.99
2025-03-18 21:51:54,109 - INFO - Class '9': TPR=0.91, TNR=0.99
Evaluating class metrics: 100%|██████████| 10/10 [00:00<00:00, 2207.53it/s]
2025-03-18 21:51:54,113 - INFO - Built-in Confusion Matrix:
[[ 953    0    2    0    2    0   14    1    8    0]
 [   1 1107    3    0    2    0    5    0   17    0]
 [  25   74  838    2   21    0   24    5   43    0]
 [  37   59   79  692   12    0   29   16   82    4]
 [   1   28    6    0  913    0   10    1   11   12]
 [  79   31   20   38   71  359   63   14  201   16]
 [  22   12    5    0   20    3  889    0    7    0]
 [  12   66   38    0   48    0    2  818   12   32]
 [  25   62    8    2   31    2   20    1  819    4]
 [  28   29    8    4  185    0    2   31   35  687]]
2025-03-18 21:51:54,114 - INFO - Overall Accuracy: 80.75%
2025-03-18 21:51:54,116 - INFO - Class '0': TPR=0.97, TNR=0.97
2025-03-18 21:51:54,116 - INFO - Class '1': TPR=0.98, TNR=0.96
2025-03-18 21:51:54,116 - INFO - Class '2': TPR=0.81, TNR=0.98
2025-03-18 21:51:54,117 - INFO - Class '3': TPR=0.69, TNR=0.99
2025-03-18 21:51:54,117 - INFO - Class '4': TPR=0.93, TNR=0.96
2025-03-18 21:51:54,117 - INFO - Class '5': TPR=0.40, TNR=1.00
2025-03-18 21:51:54,118 - INFO - Class '6': TPR=0.93, TNR=0.98
2025-03-18 21:51:54,119 - INFO - Class '7': TPR=0.80, TNR=0.99
2025-03-18 21:51:54,119 - INFO - Class '8': TPR=0.84, TNR=0.95
2025-03-18 21:51:54,120 - INFO - Class '9': TPR=0.68, TNR=0.99
Evaluating class metrics: 100%|██████████| 10/10 [00:00<00:00, 2353.31it/s]
2025-03-18 21:51:54,126 - INFO - Built-in Confusion Matrix:
[[ 954    0    1    2    2    9    7    3    2    0]
 [   0 1109    8    2    0    1    3    2   10    0]
 [   7    9  923   17    9    4   13    8   38    4]
 [   3    1   17  924    2   22    2   11   21    7]
 [   1    3    7    4  921    0    6    6    7   27]
 [   7    4    3   32    8  777   15   10   31    5]
 [  10    3    7    2    6   14  913    1    2    0]
 [   1    7   25    6    4    2    0  950    3   30]
 [   7   13    5   19    8   26    9   11  865   11]
 [  10    8    1    9   19    4    0   25    9  924]]
2025-03-18 21:51:54,126 - INFO - Overall Accuracy: 92.60%
2025-03-18 21:51:54,129 - INFO - Class '0': TPR=0.97, TNR=0.99
2025-03-18 21:51:54,129 - INFO - Class '1': TPR=0.98, TNR=0.99
2025-03-18 21:51:54,130 - INFO - Class '2': TPR=0.89, TNR=0.99
2025-03-18 21:51:54,130 - INFO - Class '3': TPR=0.91, TNR=0.99
2025-03-18 21:51:54,130 - INFO - Class '4': TPR=0.94, TNR=0.99
2025-03-18 21:51:54,131 - INFO - Class '5': TPR=0.87, TNR=0.99
2025-03-18 21:51:54,131 - INFO - Class '6': TPR=0.95, TNR=0.99
2025-03-18 21:51:54,131 - INFO - Class '7': TPR=0.92, TNR=0.99
2025-03-18 21:51:54,132 - INFO - Class '8': TPR=0.89, TNR=0.99
2025-03-18 21:51:54,132 - INFO - Class '9': TPR=0.92, TNR=0.99
Evaluating class metrics: 100%|██████████| 10/10 [00:00<00:00, 2462.17it/s]
```

```
2025-03-18 21:51:54,138 - INFO - Built-in Confusion Matrix:
[[ 958    0    1    0    1    4    5    1    9    1]
 [   0 1112    2    1    2    0    4    0   13    1]
 [  31   99  763   13   25    0   22   17   57    5]
 [  20   43   18  816    7    6    8   20   50   22]
 [   1   26    5    0  887    1    6    1   10   45]
 [  43   27    5   64   37  562   19   16   91   28]
 [  36   16    7    0   28   13  841    0   17    0]
 [   8   54    8    2   28    0    1  860    6   61]
 [  23   64    5   10   26   10   12    7  798   19]
 [  22   15    1    9   74    0    1   40   11  836]]
2025-03-18 21:51:54,138 - INFO - Overall Accuracy: 84.33%
2025-03-18 21:51:54,140 - INFO - Class '0': TPR=0.98, TNR=0.98
2025-03-18 21:51:54,141 - INFO - Class '1': TPR=0.98, TNR=0.96
2025-03-18 21:51:54,142 - INFO - Class '2': TPR=0.74, TNR=0.99
2025-03-18 21:51:54,142 - INFO - Class '3': TPR=0.81, TNR=0.99
2025-03-18 21:51:54,142 - INFO - Class '4': TPR=0.90, TNR=0.97
2025-03-18 21:51:54,143 - INFO - Class '5': TPR=0.63, TNR=1.00
2025-03-18 21:51:54,143 - INFO - Class '6': TPR=0.88, TNR=0.99
2025-03-18 21:51:54,143 - INFO - Class '7': TPR=0.84, TNR=0.99
2025-03-18 21:51:54,144 - INFO - Class '8': TPR=0.82, TNR=0.97
2025-03-18 21:51:54,144 - INFO - Class '9': TPR=0.83, TNR=0.98
Evaluating class metrics: 100%|██████████| 10/10 [00:00<00:00, 2143.89it/s]
Evaluate Regressions: 100%|██████████| 3/3 [00:00<00:00, 40.26it/s]
2025-03-18 21:51:54,146 - INFO - Evaluation complete for Perceptrons & Regressions.
```

# Visualize (Generate Plots, Confusion Matricies, etc.)

In [8]:

```
# import pandas as pd
# import seaborn as sns
# import matplotlib.pyplot as plt
# from tqdm import tqdm

################################################
# 1) CREATE A SINGLE PANDAS DATAFRAME FOR ALL RESULTS
################################################
all_rows = []

# (A) Clean PLA
for i, max_iter in tqdm(
    enumerate(perceptron_max_iter_values),
    desc="Collecting Clean PLA",
    total=len(perceptron_max_iter_values)
):
    all_rows.append({
        'model': 'Clean PLA',
        'max_iter': max_iter,
        'runtime': runtimes_clean[i],
        'accuracy': accuracies_clean[i],
        'sensitivity': sensitivities_clean[i],
        'selectivity': selectivities_clean[i]
    })

# (B) Pocket PLA
for i, max_iter in tqdm(
    enumerate(perceptron_max_iter_values),
    desc="Collecting Pocket PLA",
    total=len(perceptron_max_iter_values)
):
    all_rows.append({
        'model': 'Pocket PLA',
        'max_iter': max_iter,
        'runtime': runtimes_pocket[i],
        'accuracy': accuracies_pocket[i],
        'sensitivity': sensitivities_pocket[i],
        'selectivity': selectivities_pocket[i]
    })
```

```
# (C) Softmax
for i, row_meta in tqdm(
    enumerate(meta_soft),
    desc="Collecting Softmax",
    total=len(meta_soft)
):
    all_rows.append({
        'model': 'Softmax',
        'max_iter': row_meta['max_iter'],
        'runtime': runtimes_softmax[i],
        'accuracy': accuracies_softmax[i],
        'sensitivity': sensitivities_soft[i],
        'selectivity': selectivities_soft[i]
    })

# (D) Linear
for i, row_meta in tqdm(
    enumerate(meta_linear),
    desc="Collecting Linear",
    total=len(meta_linear)
):
    all_rows.append({
        'model': 'Linear',
        'max_iter': row_meta['max_iter'],
        'runtime': runtimes_linear[i],
        'accuracy': accuracies_linear[i],
        'sensitivity': sensitivities_lin[i],
        'selectivity': selectivities_lin[i]
    })

df_results = pd.DataFrame(all_rows)
logger.info("Combined Results DataFrame:\n%s", df_results)
display(df_results.head(20))

##############################################################################
# 2) CONFUSION MATRICES FOR ALL MODELS (GROUPED BY PLOT TYPE)
##############################################################################

logger.info("=== Plotting ALL Confusion Matrices ===")

# 2A) Perceptron: Clean
for idx, meta in tqdm(enumerate(meta_clean), total=len(meta_clean), desc="Confusions: Cle
an PLA"):
    title = f"Clean PLA (max_iter={meta['max_iter']}, Acc={meta['accuracy']*100:.2f}%)"
    plot_confusion_matrix_annotated(
        conf_clean[idx],
        classes=range(10),
        title=title,
        method=meta["method"],
        max_iter=meta["max_iter"]
    )

# 2B) Perceptron: Pocket
for idx, meta in tqdm(enumerate(meta_pocket), total=len(meta_pocket), desc="Confusions:
Pocket PLA"):
    title = f"Pocket PLA (max_iter={meta['max_iter']}, Acc={meta['accuracy']*100:.2f}%)"
    plot_confusion_matrix_annotated(
        conf_pocket[idx],
        classes=range(10),
        title=title,
        method=meta["method"],
        max_iter=meta["max_iter"]
    )

# 2C) Softmax
for idx, meta in tqdm(enumerate(meta_soft), total=len(meta_soft), desc="Confusions: Softm
ax"):
    title = f"Softmax ({meta['label']}, Acc={meta['accuracy']*100:.2f}%)"
    plot_confusion_matrix_annotated(
        conf_soft[idx],
        classes=range(10),
        title=title,
```

```
        method=meta["method"],
        max_iter=meta["max_iter"]
    )

# 2D) Linear
for idx, meta in tqdm(enumerate(meta_linear), total=len(meta_linear), desc="Confusions:
Linear"):
    title = f"Linear ({meta['label']}, Acc={meta['accuracy']*100:.2f}%)"
    plot_confusion_matrix_annotated(
        conf_linear[idx],
        classes=range(10),
        title=title,
        method=meta["method"],
        max_iter=meta["max_iter"]
    )


##############################################################################
# 3) ITERATION-LEVEL PLOTS (ALL MODELS)
##############################################################################

logger.info("=== Iteration-Level Visualization (All Models) ===")

# 3A) Perceptron: Clean & Pocket
for max_iter, c_model in trained_models_clean.items():
    df_iter = c_model.get_iteration_df()
    if not df_iter.empty and "train_error" in df_iter.columns:
        title = f"Clean PLA max_iter={max_iter}: Train Error vs. Iteration"
        df_iter.plot(x="iteration", y="train_error", marker='o', figsize=(8,5), title=ti
tle)
        plt.grid(True, linestyle='--', alpha=0.7)
        plt.show()

for max_iter, p_model in trained_models_pocket.items():
    df_iter = p_model.get_iteration_df()
    if not df_iter.empty and "train_error" in df_iter.columns:
        title = f"Pocket PLA max_iter={max_iter}: Train Error vs. Iteration"
        df_iter.plot(x="iteration", y="train_error", marker='o', figsize=(8,5), title=ti
tle)
        plt.grid(True, linestyle='--', alpha=0.7)
        plt.show()

# 3B) Softmax
for (lr_val, max_iter_val), s_model in trained_models_softmax.items():
    df_iter = s_model.get_iteration_df()  # Must be implemented in your SoftmaxRegressio
n
    if not df_iter.empty:
        title = f"Softmax LR={lr_val}, max_iter={max_iter_val}: Train Loss vs. Iteration"
        df_iter.plot(x="iteration", y="train_loss", marker='o', figsize=(8,5), title=tit
le)
        plt.grid(True, linestyle='--', alpha=0.7)
        plt.show()

        if "test_loss" in df_iter.columns:
            title = f"Softmax LR={lr_val}, max_iter={max_iter_val}: Train & Test Loss"
            df_iter.plot(x="iteration", y=["train_loss","test_loss"], marker='o', figsiz
e=(8,5), title=title)
            plt.grid(True, linestyle='--', alpha=0.7)
            plt.show()

        if "avg_adaptive_lr" in df_iter.columns:
            title = f"Softmax LR={lr_val}, max_iter={max_iter_val}: Avg Adaptive LR vs. I
teration"
            df_iter.plot(x="iteration", y="avg_adaptive_lr", marker='x', figsize=(8,5),
title=title)
            plt.grid(True, linestyle='--', alpha=0.7)
            plt.show()

# 3C) Linear
for (lr_val, max_iter_val), lin_model in trained_models_linear.items():
    df_iter = lin_model.get_iteration_df()  # Must be implemented in your LinearRegressi
on
```

```
    if not df_iter.empty:
        title = f"Linear LR={lr_val}, max_iter={max_iter_val}: Train Loss vs. Iteration"
        df_iter.plot(x="iteration", y="train_loss", marker='o', figsize=(8,5), title=tit
le)
        plt.grid(True, linestyle='--', alpha=0.7)
        plt.show()

        if "test_loss" in df_iter.columns:
            title = f"Linear LR={lr_val}, max_iter={max_iter_val}: Train & Test Loss"
            df_iter.plot(x="iteration", y=["train_loss","test_loss"], marker='o', figsiz
e=(8,5), title=title)
            plt.grid(True, linestyle='--', alpha=0.7)
            plt.show()

        if "avg_adaptive_lr" in df_iter.columns:
            title = f"Linear LR={lr_val}, max_iter={max_iter_val}: Avg Adaptive LR vs. It
eration"
            df_iter.plot(x="iteration", y="avg_adaptive_lr", marker='x', figsize=(8,5),
title=title)
            plt.grid(True, linestyle='--', alpha=0.7)
            plt.show()


##############################################################################
# 4) PANDAS + SEABORN PLOTS
##############################################################################

logger.info("=== Pandas + Seaborn Plots ===")

# 4A) LINE PLOT: Accuracy vs. max_iter (Perceptrons Only)
df_perc = df_results[df_results['model'].isin(['Clean PLA','Pocket PLA'])].copy()
df_perc.sort_values(['model','max_iter'], inplace=True)

plt.figure(figsize=(6,4))
sns.lineplot(
    data=df_perc,
    x='max_iter', y='accuracy',
    hue='model', marker='o'
)
plt.title("Perceptrons: Accuracy vs. max_iter (Pandas/Seaborn)")
plt.grid(True, linestyle='--', alpha=0.7)
plt.show()

# 4B) BAR CHART: Average Accuracy by Model
df_mean = df_results.groupby('model', as_index=False)['accuracy'].mean()

plt.figure(figsize=(6,4))
sns.barplot(data=df_mean, x='model', y='accuracy')
plt.title("Average Accuracy by Model (Pandas/Seaborn)")
plt.ylim(0.7, 1.0)
plt.grid(True, axis='y', linestyle='--', alpha=0.7)
plt.show()

# 4C) SCATTER PLOT: Accuracy vs. Runtime, colored by model
plt.figure(figsize=(6,4))
sns.scatterplot(
    data=df_results,
    x='runtime', y='accuracy',
    hue='model', style='model',
    s=100
)
plt.title("Accuracy vs. Runtime (All Models) (Pandas/Seaborn)")
plt.grid(True, linestyle='--', alpha=0.7)
plt.show()


##############################################################################
# 5) CUSTOM SUMMARY PLOTS (AGGREGATED CURVES, ETC.)
##############################################################################

logger.info("=== Custom Summaries (Aggregated Curves, etc.) ===")
```

```
# 5A) Aggregated Perceptron Curves
plot_train_curves_three_models(
    clean_train_curve=clean_train_curve,
    pocket_train_curve=pocket_train_curve,
    softmax_train_curve=None,  # no Softmax aggregator
    title="Aggregated Perceptron Train Curves (Clean vs. Pocket)",
    max_iter=perceptron_max_iter_values[-1]
)

# 5B) Summaries for Perceptron
plot_accuracy_vs_max_iter(
    max_iter_values=perceptron_max_iter_values,
    accuracies_clean=accuracies_clean,
    accuracies_pocket=accuracies_pocket,
    accuracies_softmax=None
)

plot_runtime_vs_max_iter(
    max_iter_values=perceptron_max_iter_values,
    runtimes_clean=runtimes_clean,
    runtimes_pocket=runtimes_pocket,
    runtimes_softmax=None
)

plot_accuracy_vs_runtime(
    runtimes_clean=runtimes_clean,
    accuracies_clean=accuracies_clean,
    runtimes_pocket=runtimes_pocket,
    accuracies_pocket=accuracies_pocket,
    title="Perceptrons: Accuracy vs. Runtime"
)

plot_performance_summary_extended_by_runtime(
    runtimes_clean=runtimes_clean,
    accuracies_clean=accuracies_clean,
    sensitivities_clean=sensitivities_clean,
    selectivities_clean=selectivities_clean,
    runtimes_pocket=runtimes_pocket,
    accuracies_pocket=accuracies_pocket,
    sensitivities_pocket=sensitivities_pocket,
    selectivities_pocket=selectivities_pocket,
    title="Perceptrons: Performance vs. Runtime"
)

# 5C) Summaries for Softmax & Linear
plot_accuracy_vs_runtime(
    runtimes_clean=runtimes_softmax,
    accuracies_clean=accuracies_softmax,
    title="Softmax: Accuracy vs. Runtime"
)
plot_accuracy_vs_runtime(
    runtimes_clean=runtimes_linear,
    accuracies_clean=accuracies_linear,
    title="Linear: Accuracy vs. Runtime"
)
plot_accuracy_vs_runtime(
    runtimes_clean=runtimes_softmax,
    accuracies_clean=accuracies_softmax,
    runtimes_pocket=runtimes_linear,
    accuracies_pocket=accuracies_linear,
    title="Softmax vs. Linear: Accuracy vs. Runtime"
)
plot_performance_summary_extended_by_runtime(
    runtimes_clean=runtimes_softmax,
    accuracies_clean=accuracies_softmax,
    sensitivities_clean=sensitivities_soft,
    selectivities_clean=selectivities_soft,
    runtimes_pocket=runtimes_linear,
    accuracies_pocket=accuracies_linear,
    sensitivities_pocket=sensitivities_lin,
    selectivities_pocket=selectivities_lin,
    title="Softmax vs. Linear: TPR/TNR vs. Runtime"
```

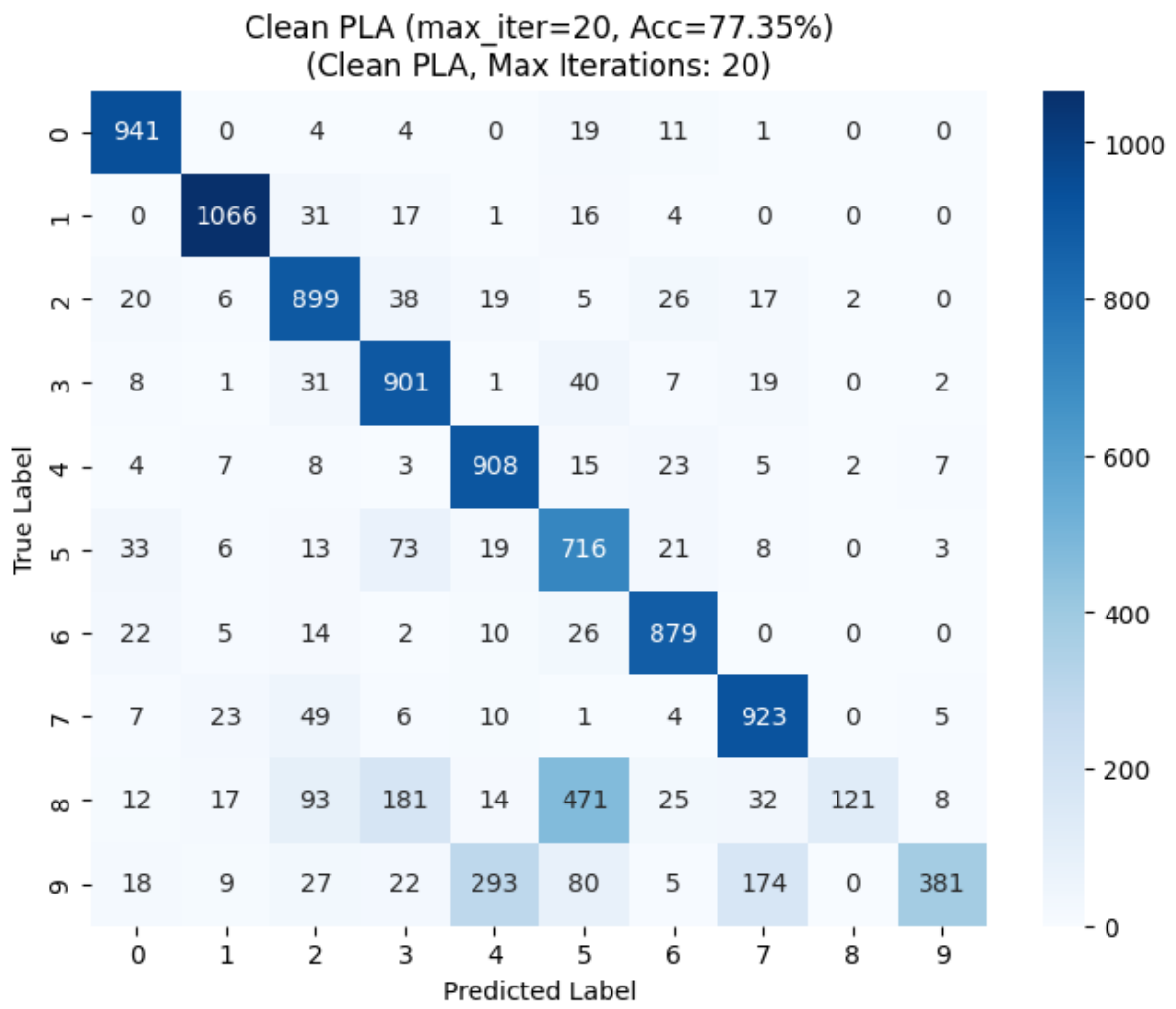
```
)

# 5D) 4-Model Comparison
plot_performance_summary_4models_by_runtime(
    runtimes_clean, accuracies_clean, sensitivities_clean, selectivities_clean,
    runtimes_pocket, accuracies_pocket, sensitivities_pocket, selectivities_pocket,
    runtimes_softmax, accuracies_softmax, sensitivities_soft, selectivities_soft,
    runtimes_linear, accuracies_linear, sensitivities_lin, selectivities_lin,
    title="Performance vs. Runtime (4-Model Comparison)"
)

plot_accuracy_vs_runtime_4models(
    rt_clean=runtimes_clean,
    acc_clean=accuracies_clean,
    rt_pocket=runtimes_pocket,
    acc_pocket=accuracies_pocket,
    rt_softmax=runtimes_softmax,
    acc_softmax=accuracies_softmax,
    rt_linear=runtimes_linear,
    acc_linear=accuracies_linear,
    title="Accuracy vs. Runtime (4 Models)"
)

logger.info("=== All Visualizations Complete ===")
```

```
Collecting Clean PLA: 100%|██████████| 4/4 [00:00<00:00, 78033.56it/s]
Collecting Pocket PLA: 100%|██████████| 4/4 [00:00<00:00, 86928.58it/s]
Collecting Softmax: 100%|██████████| 3/3 [00:00<00:00, 17003.94it/s]
Collecting Linear: 100%|██████████| 3/3 [00:00<00:00, 79137.81it/s]
2025-03-18 21:51:54,173 - INFO - Combined Results DataFrame:
         model  max_iter     runtime  accuracy  sensitivity  selectivity
0    Clean PLA        20   12.627623    0.7735     0.770718     0.974903
1    Clean PLA        50   24.711728    0.8545     0.850681     0.983826
2    Clean PLA       100   42.835001    0.8745     0.871954     0.986055
3    Clean PLA      1000  343.271393    0.8633     0.861587     0.984807
4   Pocket PLA        20   12.686805    0.8377     0.836411     0.981993
5   Pocket PLA        50   24.680696    0.8753     0.873371     0.986151
6   Pocket PLA       100   42.614644    0.8964     0.894188     0.988493
7   Pocket PLA      1000  343.552508    0.9045     0.903203     0.989398
8      Softmax       100    4.108804    0.9174     0.916210     0.990829
9      Softmax      1000   40.369971    0.9272     0.926119     0.991917
10     Softmax     10000  307.982953    0.9260     0.924975     0.991782
11      Linear       100    3.197827    0.6820     0.682591     0.964741
12      Linear      1000   32.946327    0.8075     0.802257     0.978577
13      Linear     10000  330.336316    0.8433     0.840014     0.982548
```

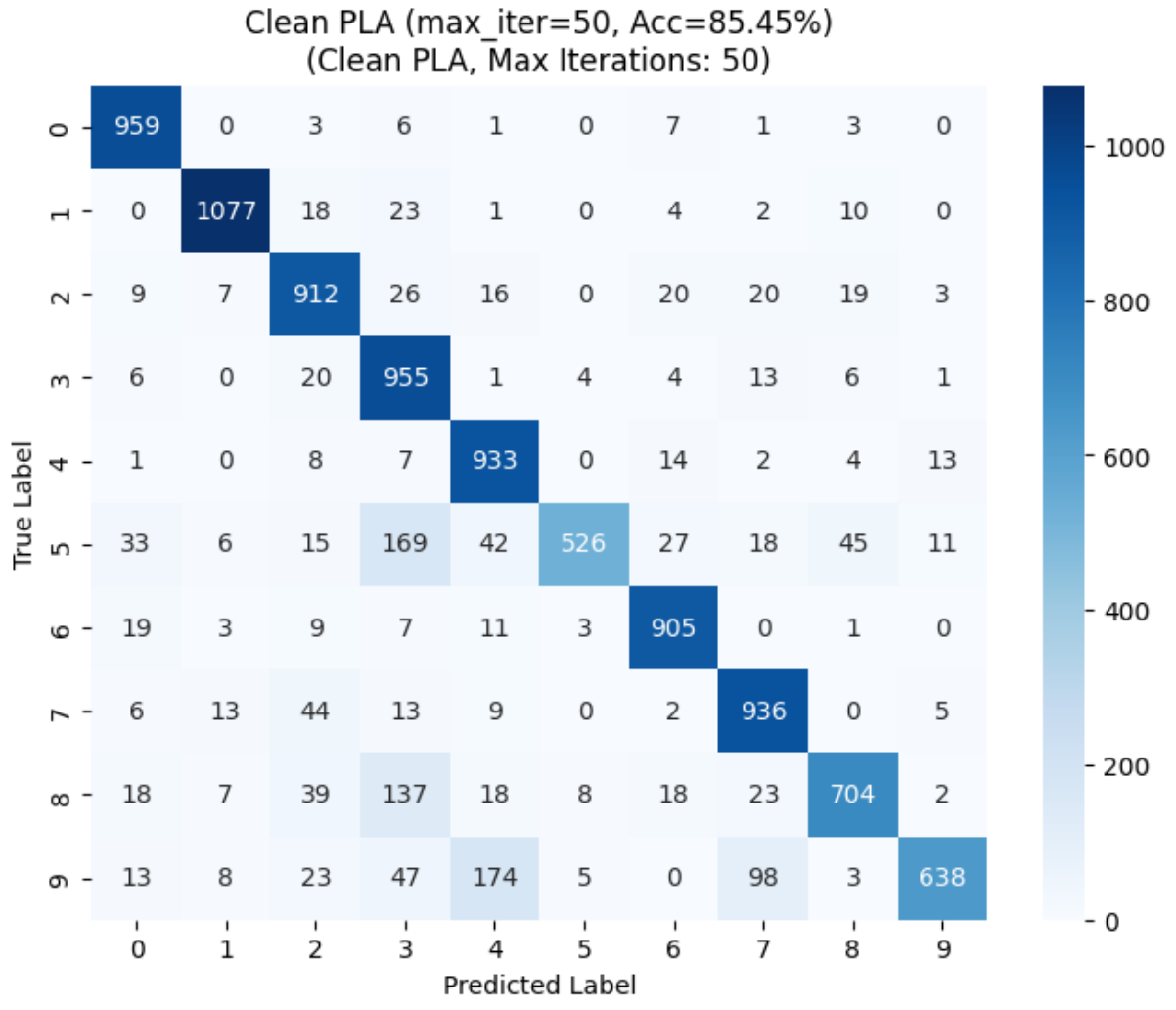
| | model | max_iter | runtime | accuracy | sensitivity | selectivity |
|---|---|---|---|---|---|---|
| **0** | **Clean PLA** | **20** | **12.627623** | **0.7735** | **0.770718** | **0.974903** |
| **1** | **Clean PLA** | **50** | **24.711728** | **0.8545** | **0.850681** | **0.983826** |
| **2** | **Clean PLA** | **100** | **42.835001** | **0.8745** | **0.871954** | **0.986055** |
| **3** | **Clean PLA** | **1000** | **343.271393** | **0.8633** | **0.861587** | **0.984807** |
| **4** | **Pocket PLA** | **20** | **12.686805** | **0.8377** | **0.836411** | **0.981993** |
| **5** | **Pocket PLA** | **50** | **24.680696** | **0.8753** | **0.873371** | **0.986151** |
| **6** | **Pocket PLA** | **100** | **42.614644** | **0.8964** | **0.894188** | **0.988493** |
| **7** | **Pocket PLA** | **1000** | **343.552508** | **0.9045** | **0.903203** | **0.989398** |
| **8** | **Softmax** | **100** | **4.108804** | **0.9174** | **0.916210** | **0.990829** |
| **9** | **Softmax** | **1000** | **40.369971** | **0.9272** | **0.926119** | **0.991917** |
| **10** | **Softmax** | **10000** | **307.982953** | **0.9260** | **0.924975** | **0.991782** |
| **11** | **Linear** | **100** | **3.197827** | **0.6820** | **0.682591** | **0.964741** |
| **12** | **Linear** | **1000** | **32.946327** | **0.8075** | **0.802257** | **0.978577** |
| **13** | **Linear** | **10000** | **330.336316** | **0.8433** | **0.840014** | **0.982548** |

```
2025-03-18 21:51:54,185 - INFO - === Plotting ALL Confusion Matrices ===
```

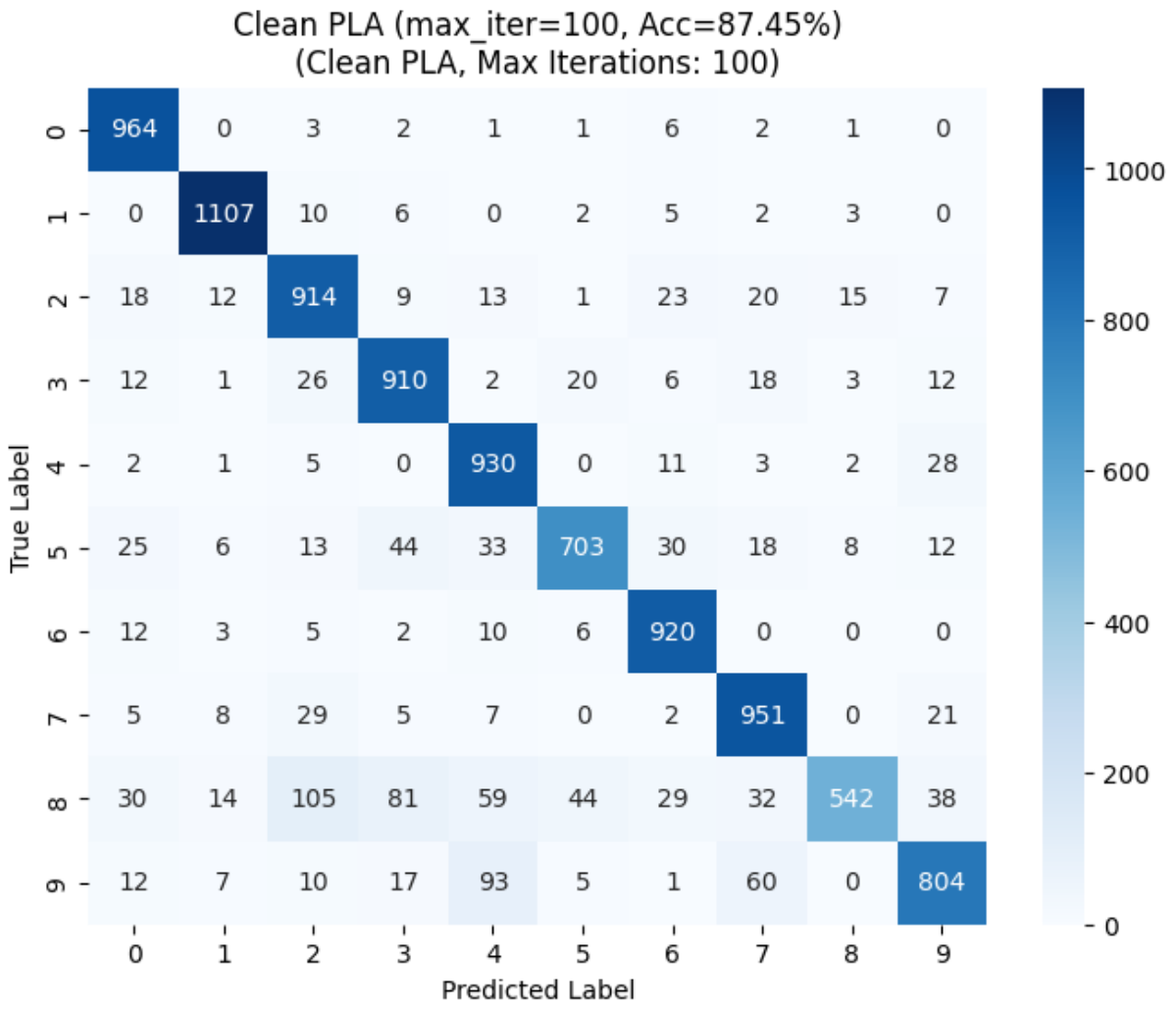
```
Confusions: Clean PLA:   0%|          | 0/4 [00:00<?, ?it/s]
```



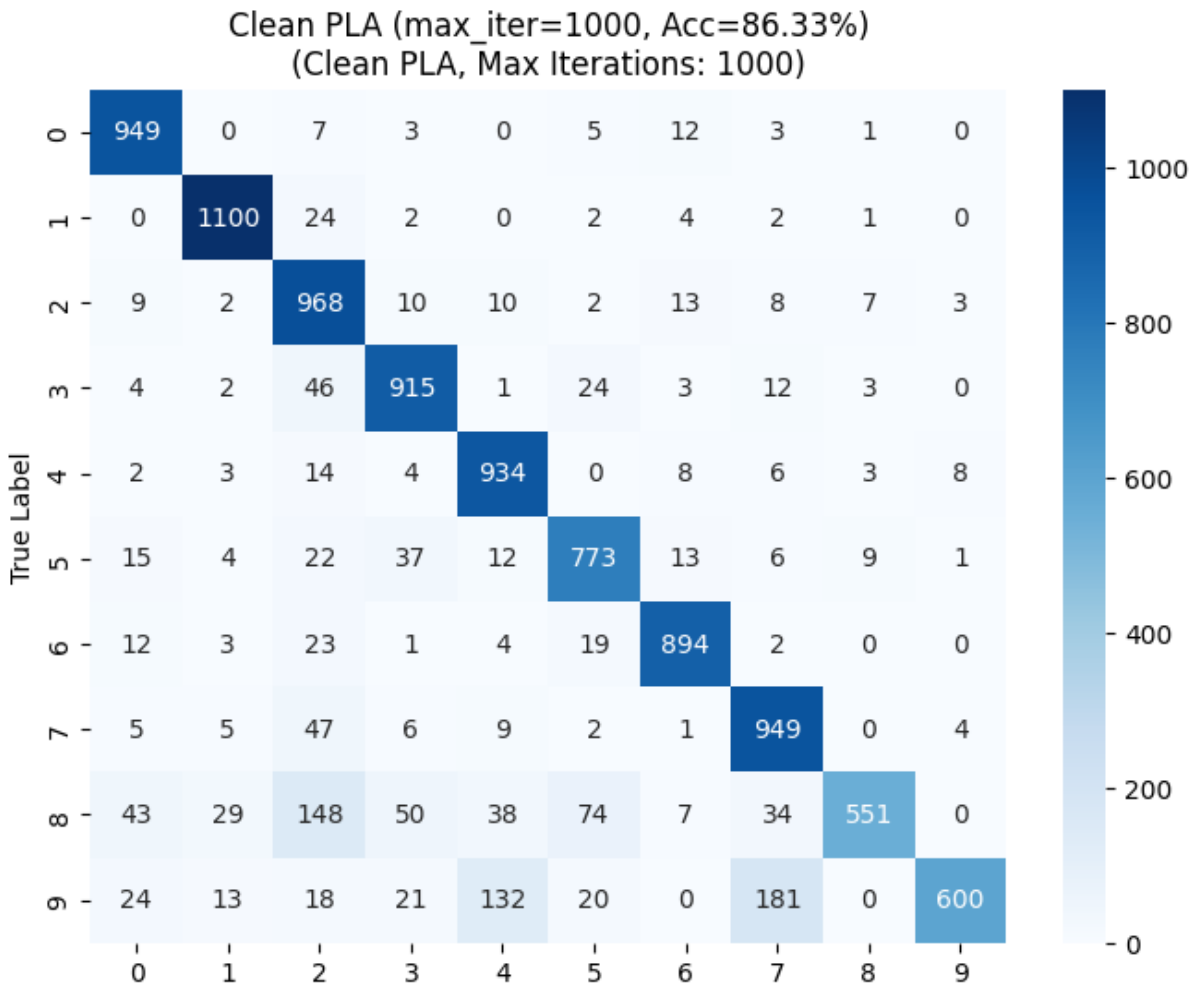

```
Confusions: Clean PLA:  25%|██▌       | 1/4 [00:00<00:00,  7.18it/s]
```



```
Confusions: Clean PLA:  50%|█████     | 2/4 [00:00<00:00,  7.06it/s]
```

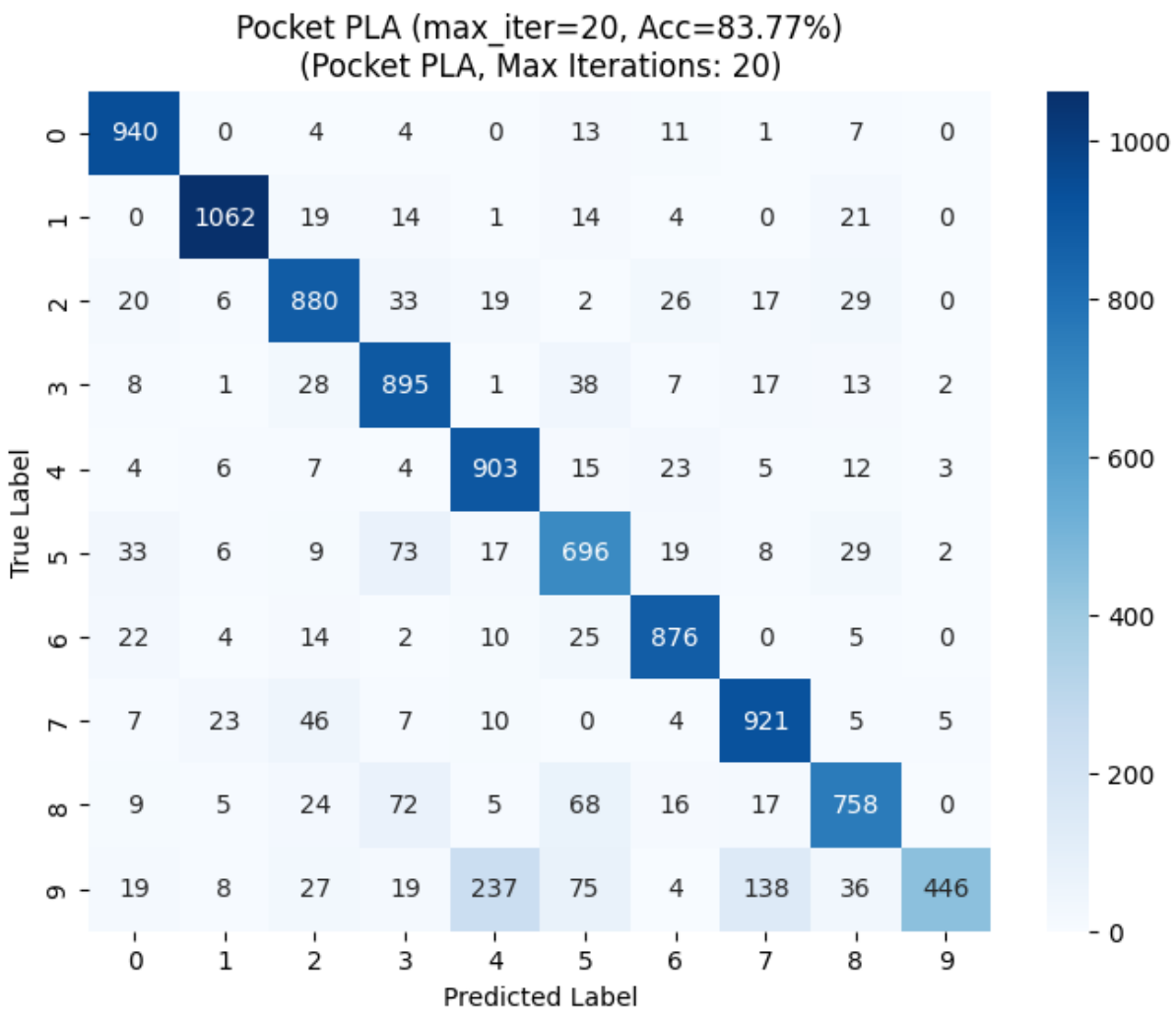



```
Confusions: Clean PLA:  75%|███████   | 3/4 [00:00<00:00,  7.28it/s]
```

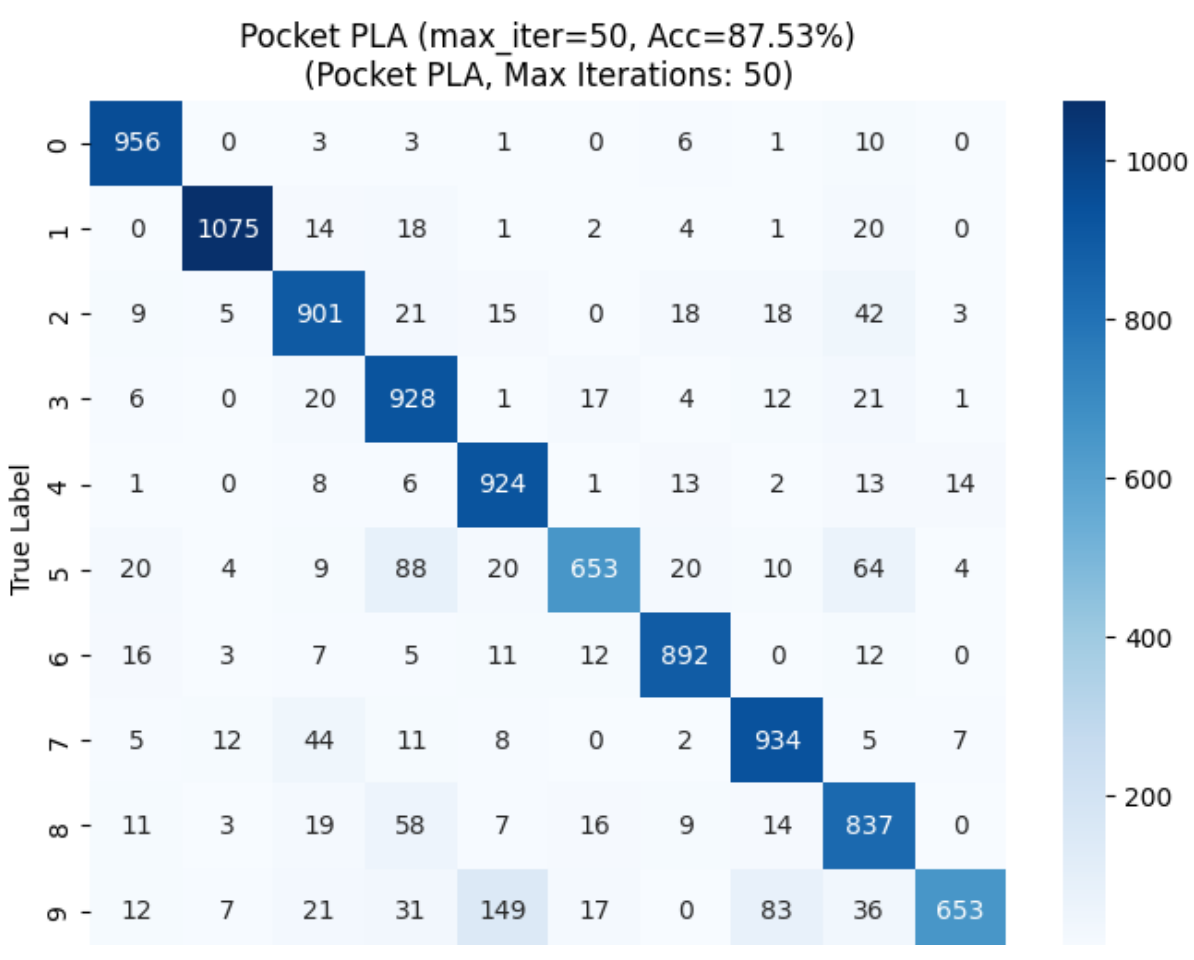


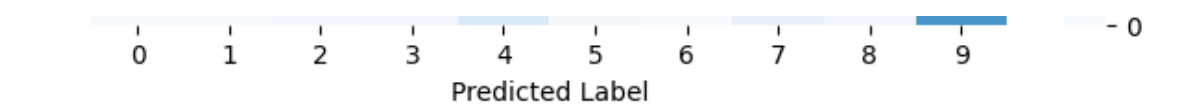
Predicted Label

```
Confusions: Clean PLA: 100%|██████████| 4/4 [00:00<00:00,  7.61it/s]
Confusions: Pocket PLA:   0%|          | 0/4 [00:00<?, ?it/s]
```

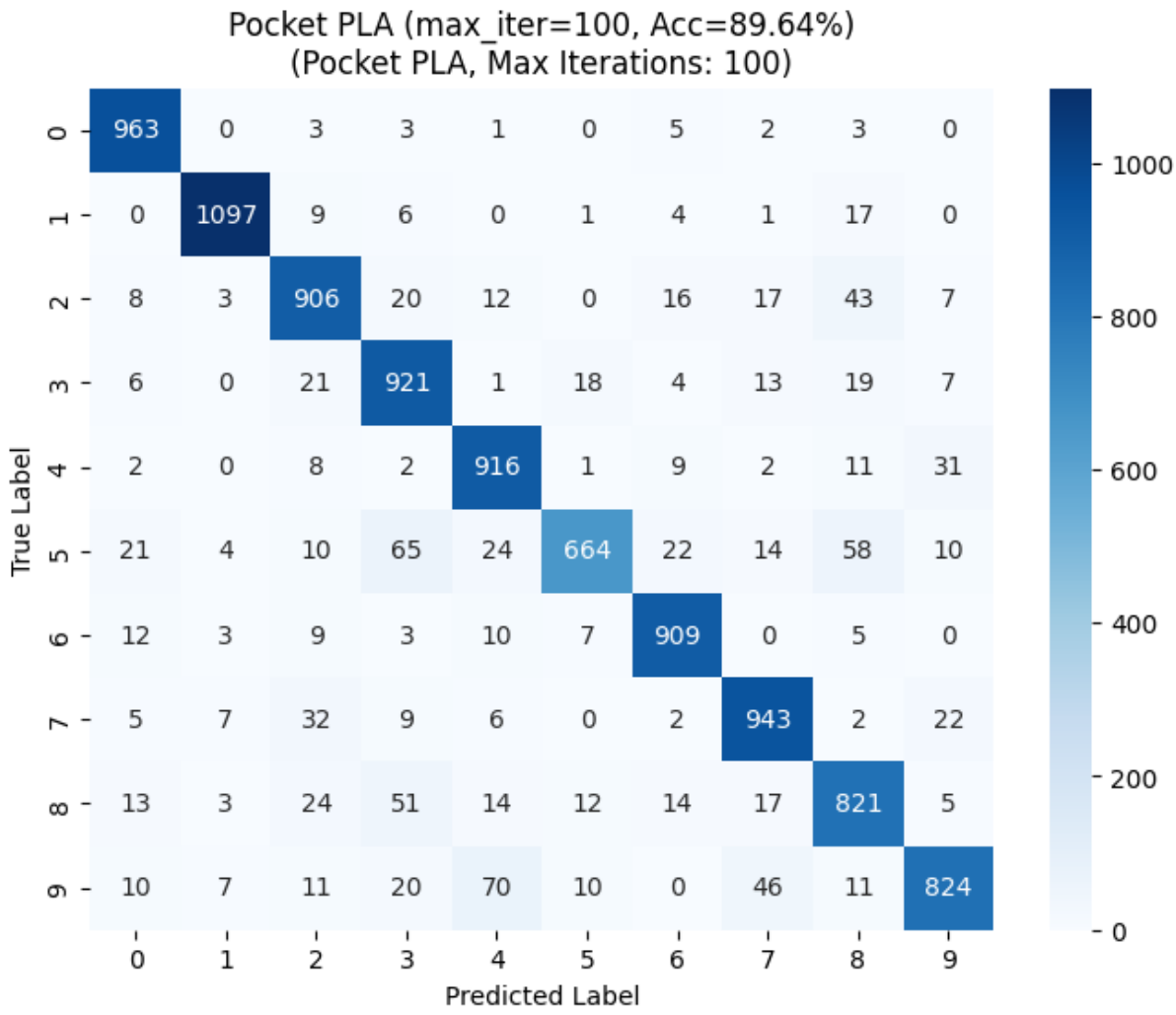



```
Confusions: Pocket PLA:  25%|███       | 1/4 [00:00<00:00,  9.36it/s]
```

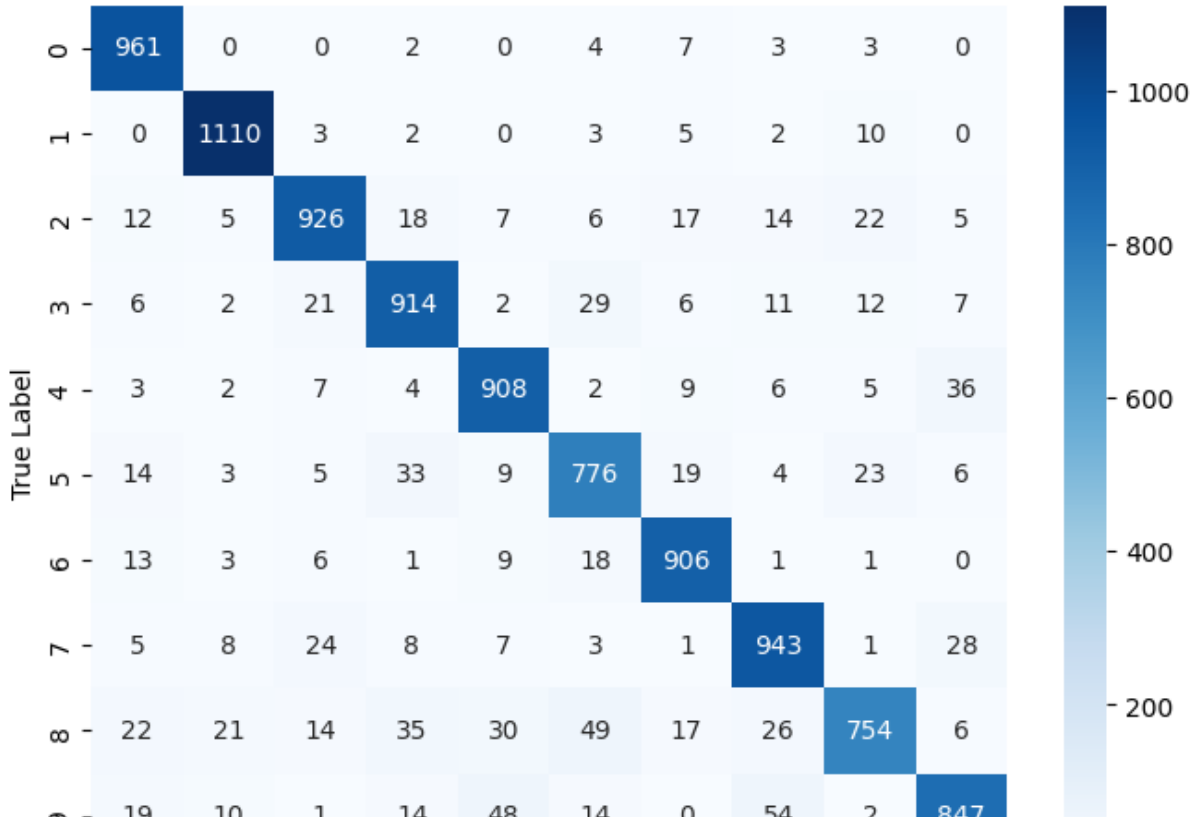


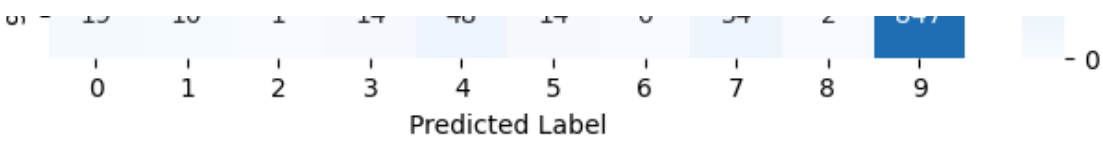



Confusions: Pocket PLA: 50%|████ | 2/4 [00:00<00:00, 9.14it/s]

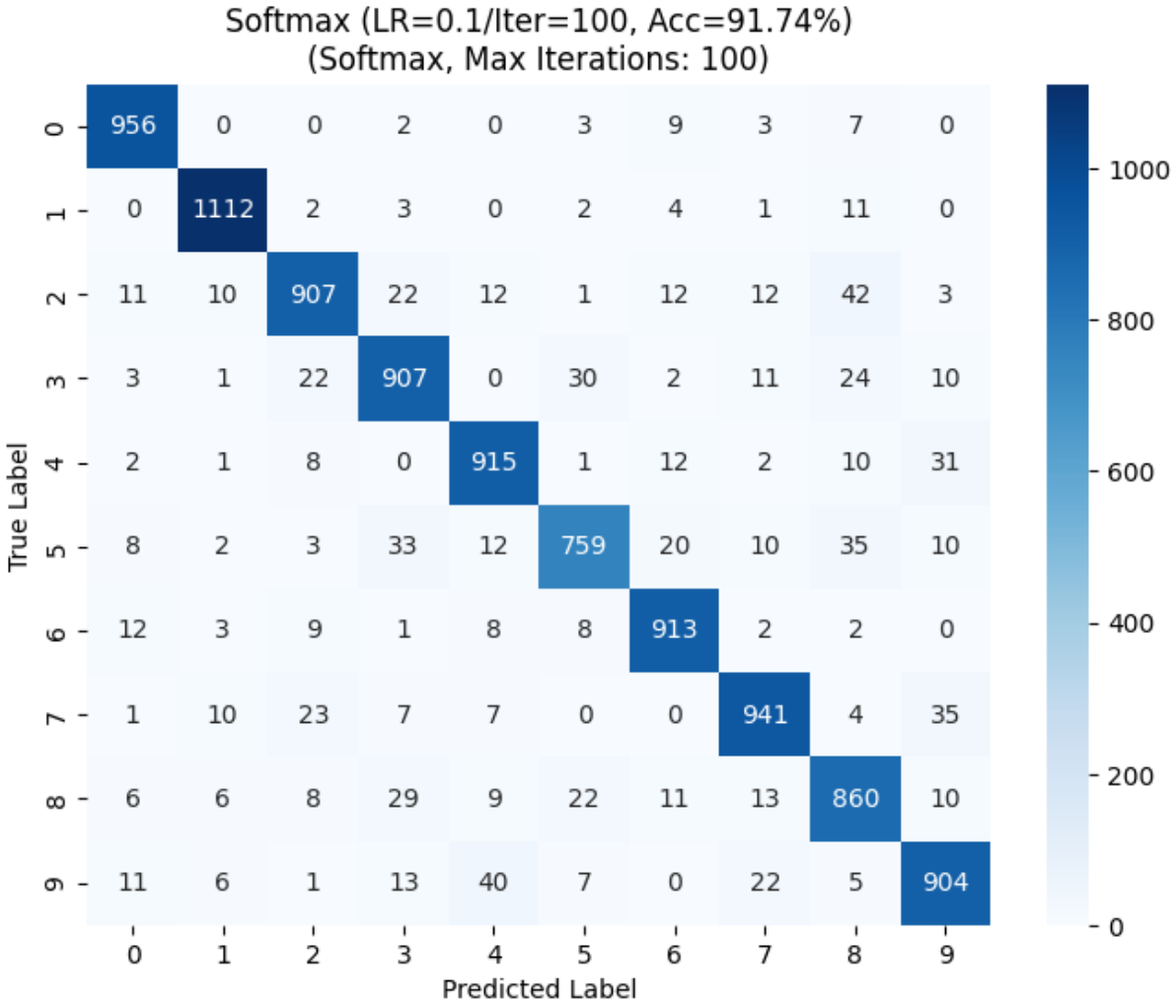



Confusions: Pocket PLA: 75%|██████ | 3/4 [00:00<00:00, 8.94it/s]

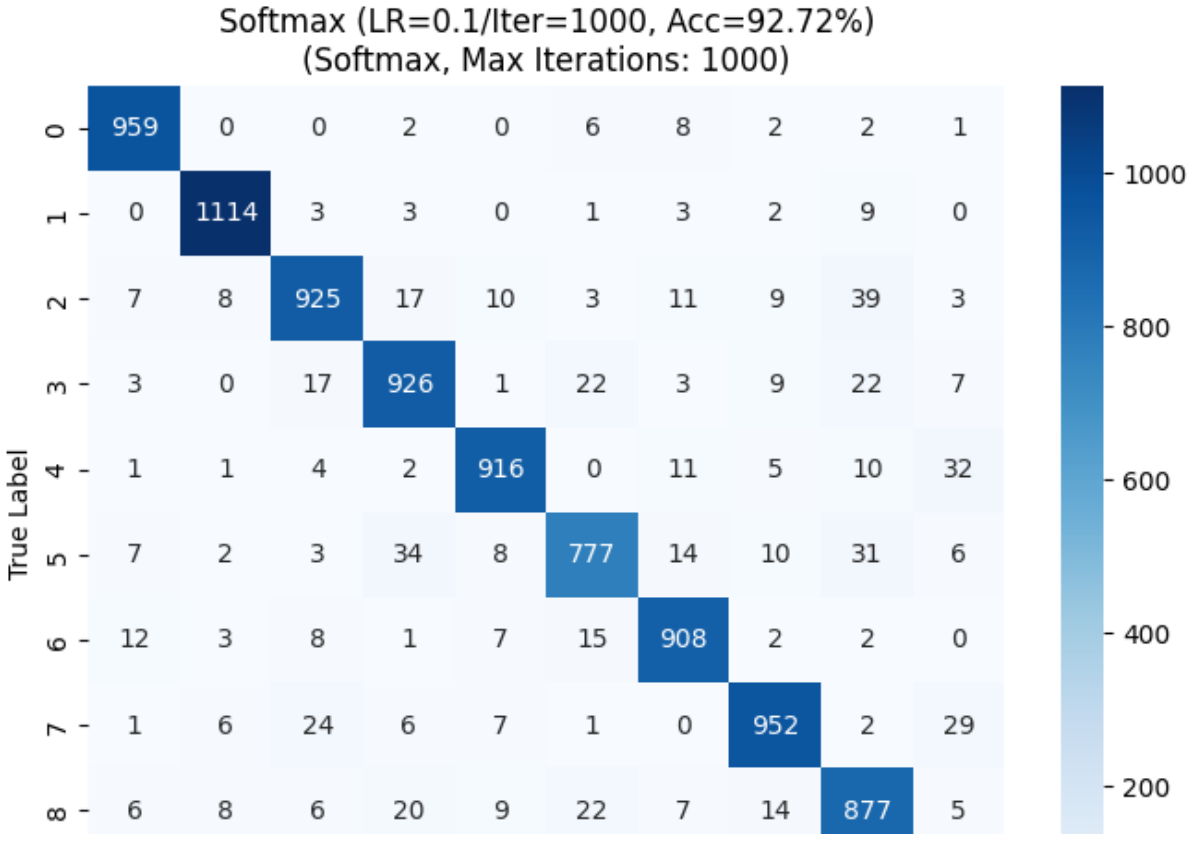


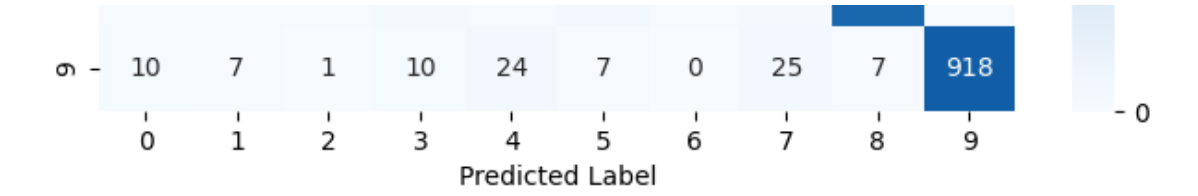



```
Confusions: Pocket PLA: 100%|██████████| 4/4 [00:00<00:00,  7.96it/s]
Confusions: Softmax:   0%|          | 0/3 [00:00<?, ?it/s]
```

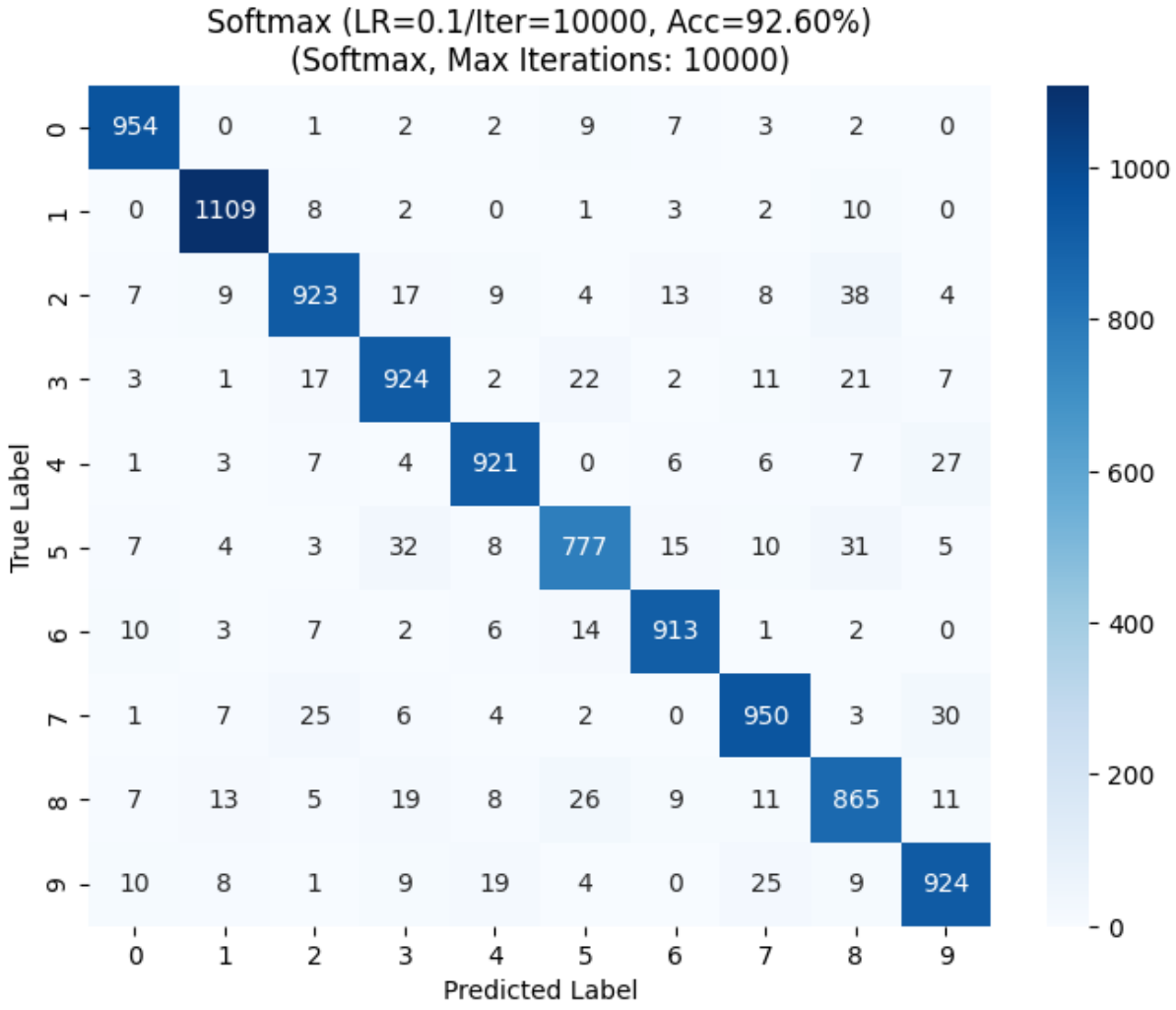



```
Confusions: Softmax:  33%|███       | 1/3 [00:00<00:00,  9.53it/s]
```

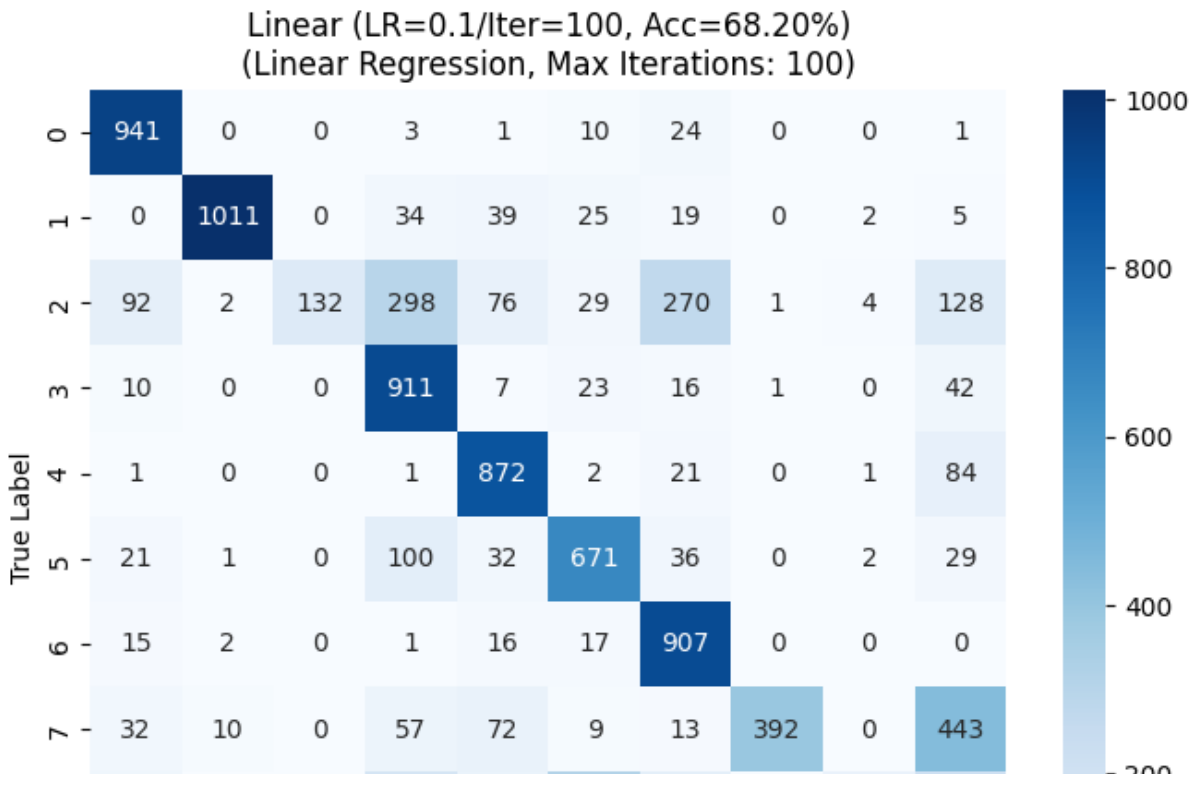


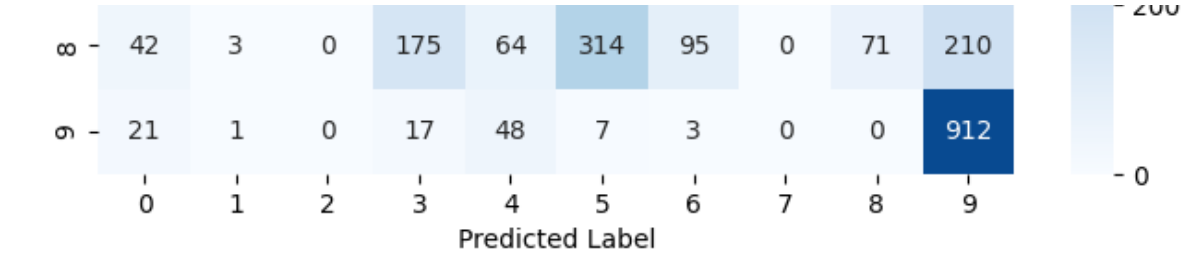



```
Confusions: Softmax:  67%|██████    | 2/3 [00:00<00:00,  9.38it/s]
```

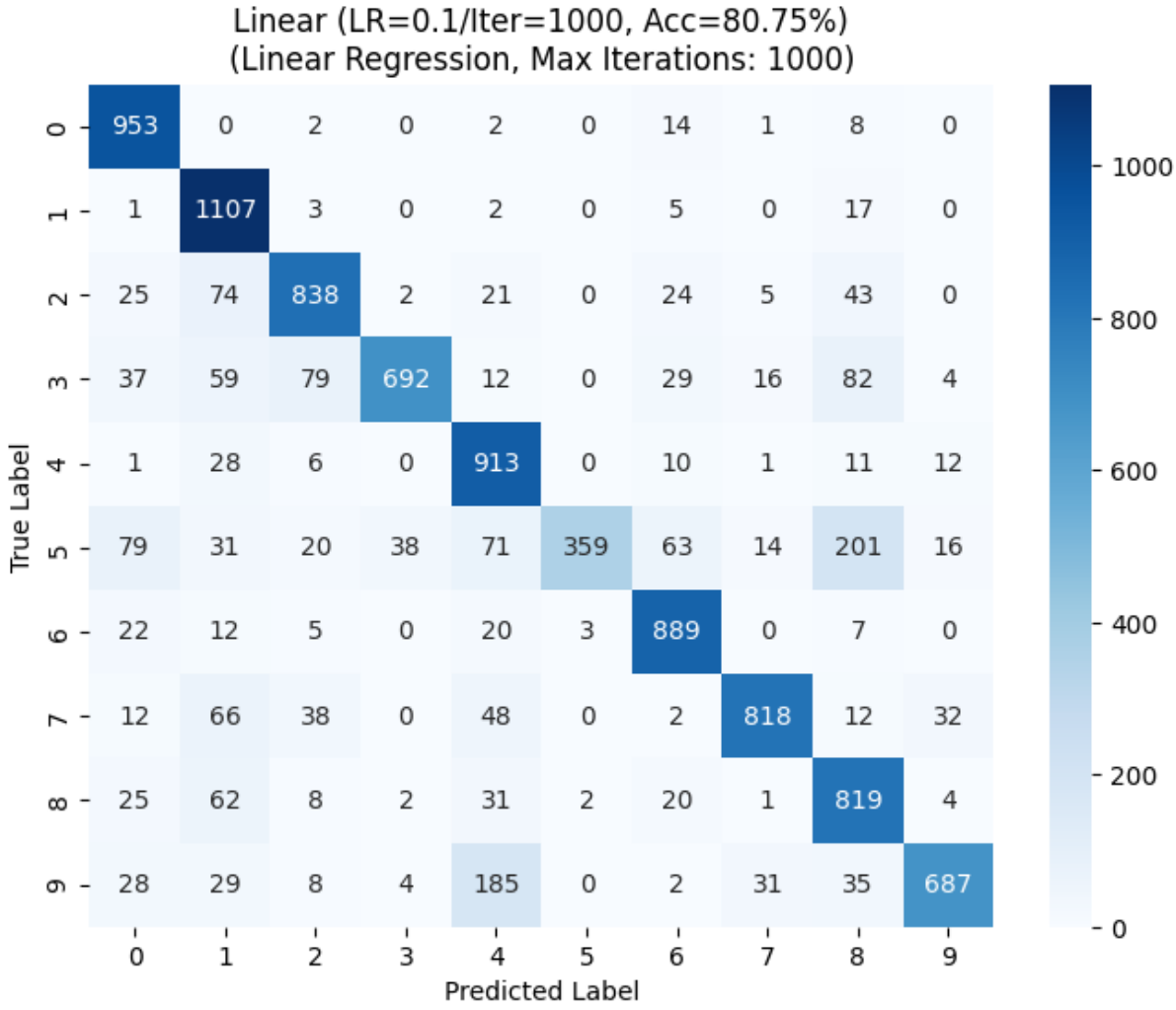



```
Confusions: Softmax: 100%|██████████| 3/3 [00:00<00:00,  9.22it/s]
Confusions: Linear:   0%|          | 0/3 [00:00<?, ?it/s]
```

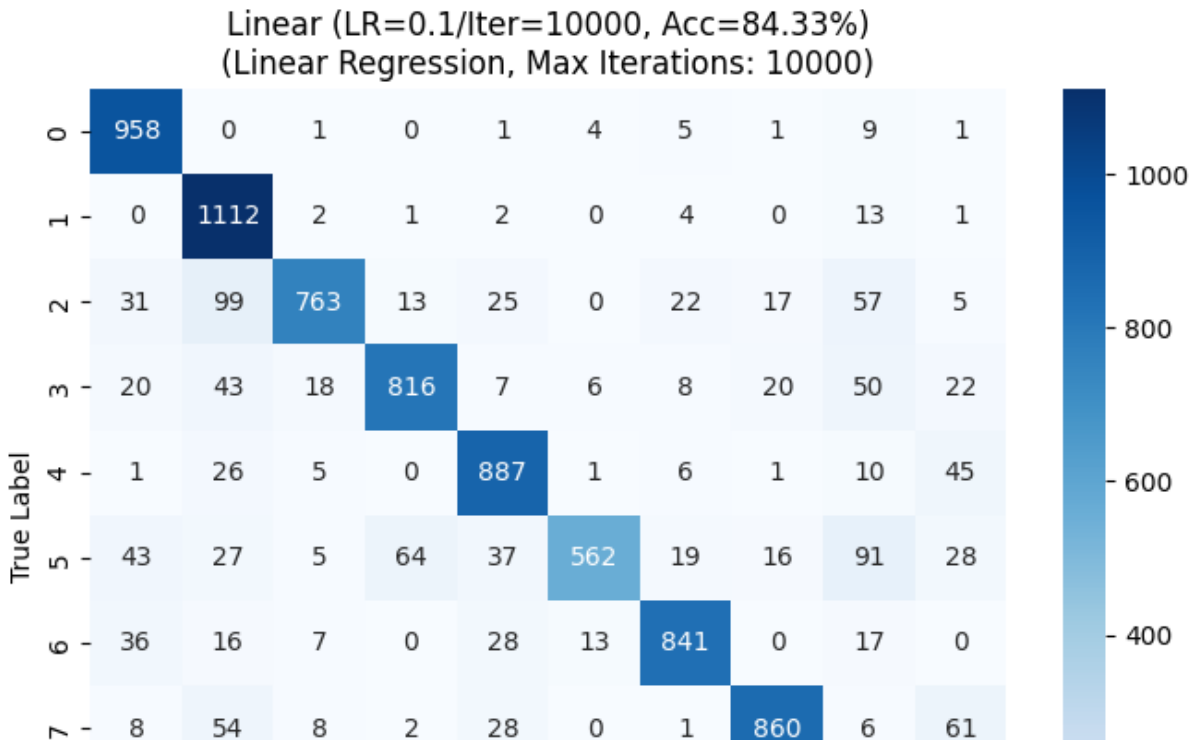


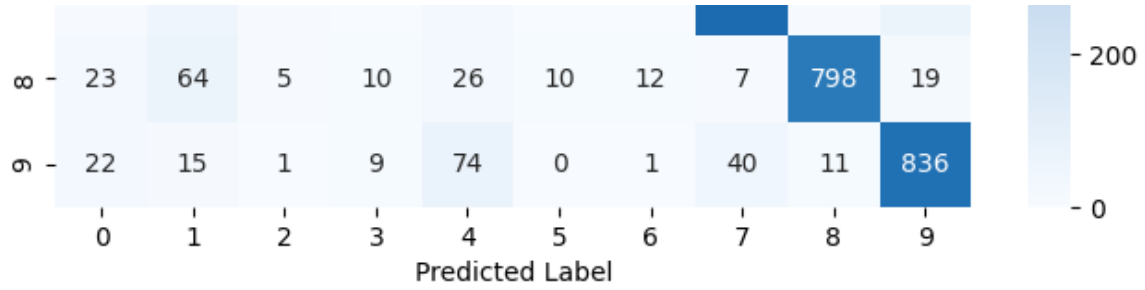



```
Confusions: Linear:  33%|███       | 1/3 [00:00<00:00,  8.65it/s]
```

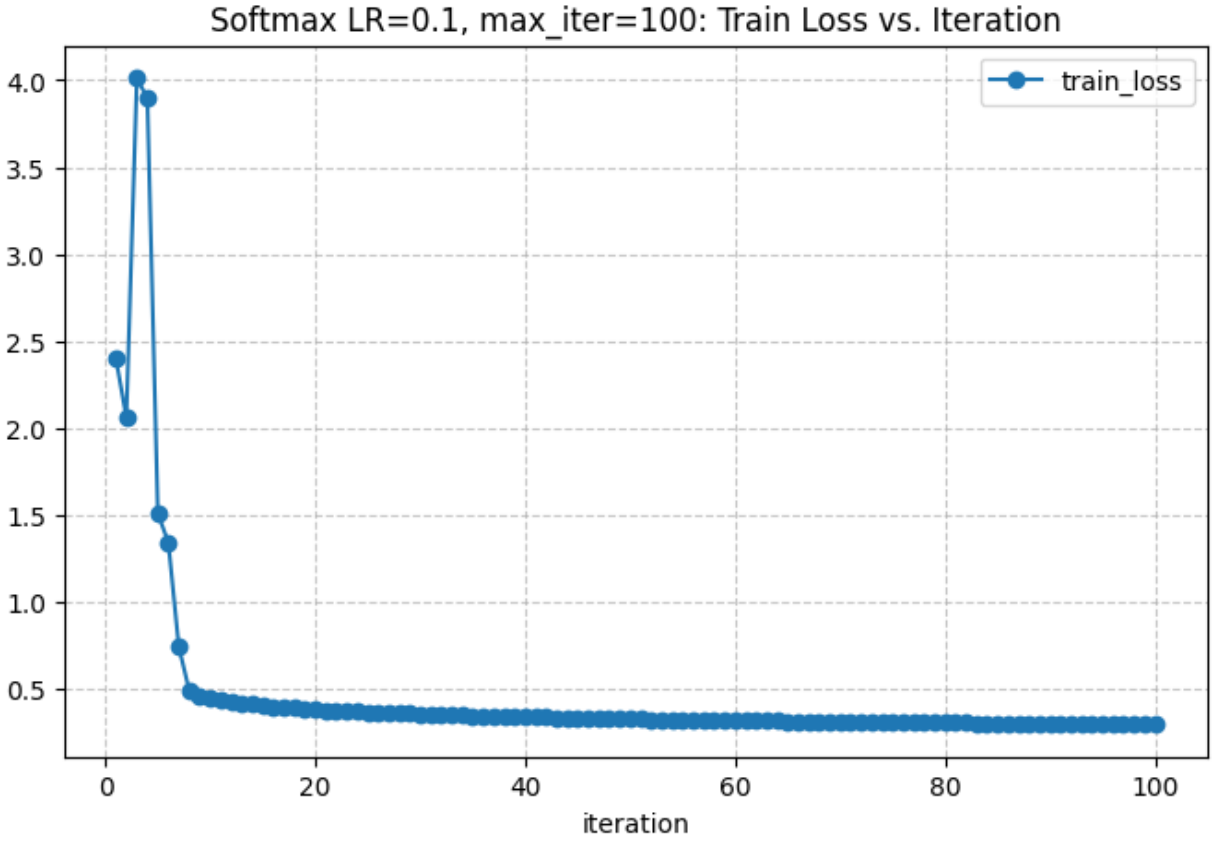


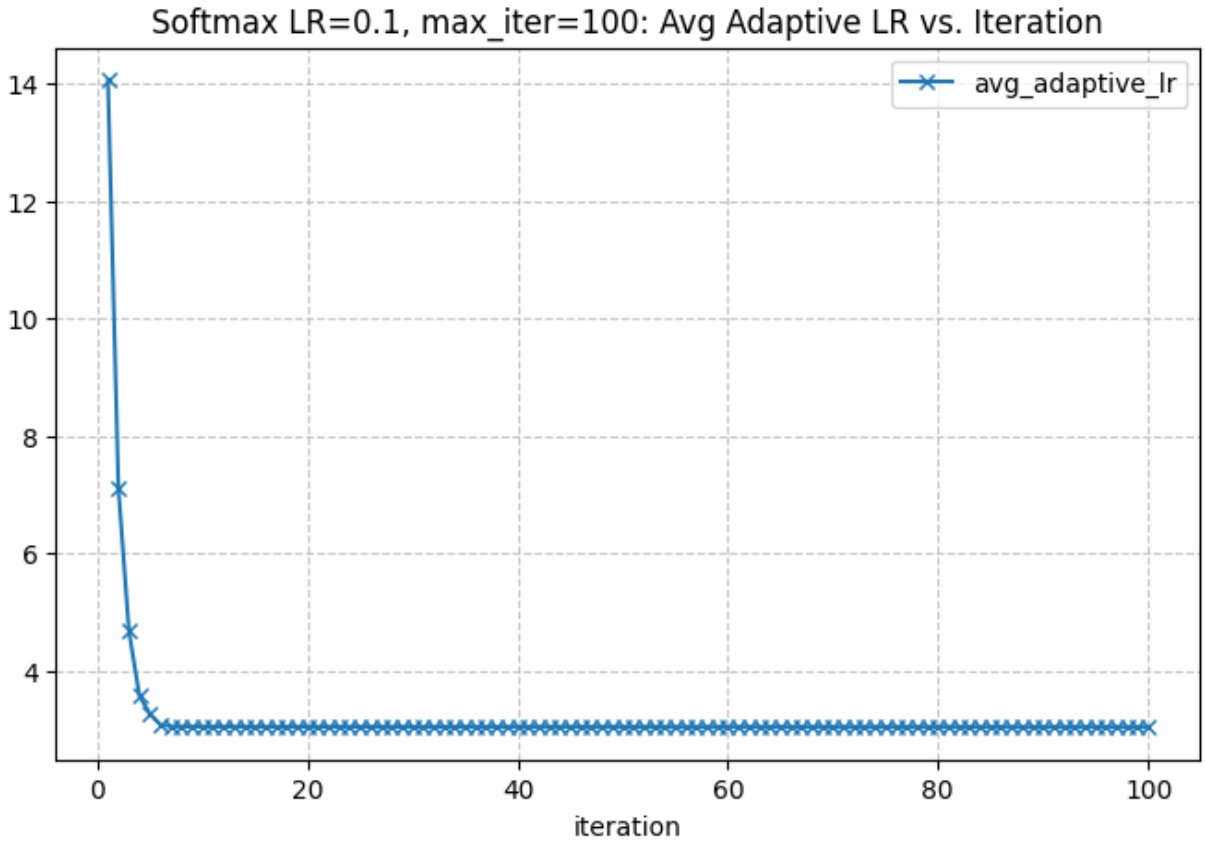

```
Confusions: Linear:  67%|██████▋   | 2/3 [00:00<00:00,  8.90it/s]
```



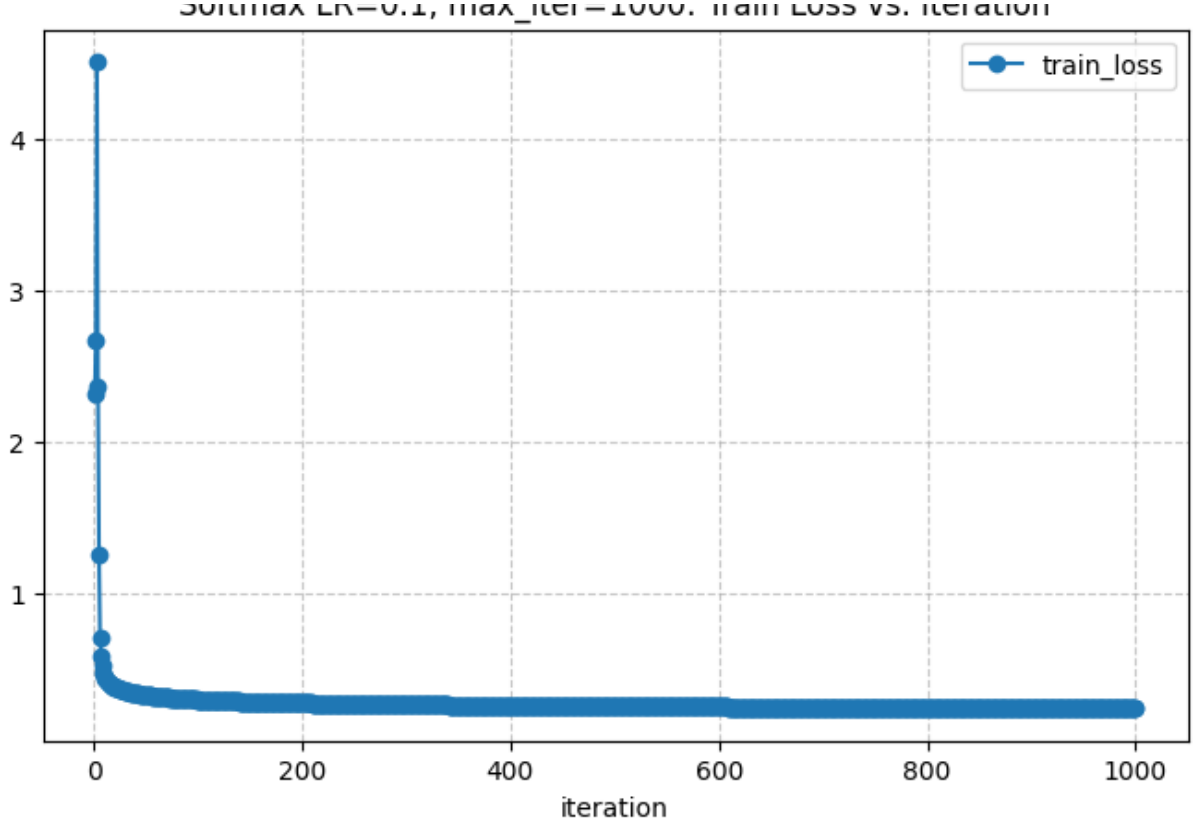


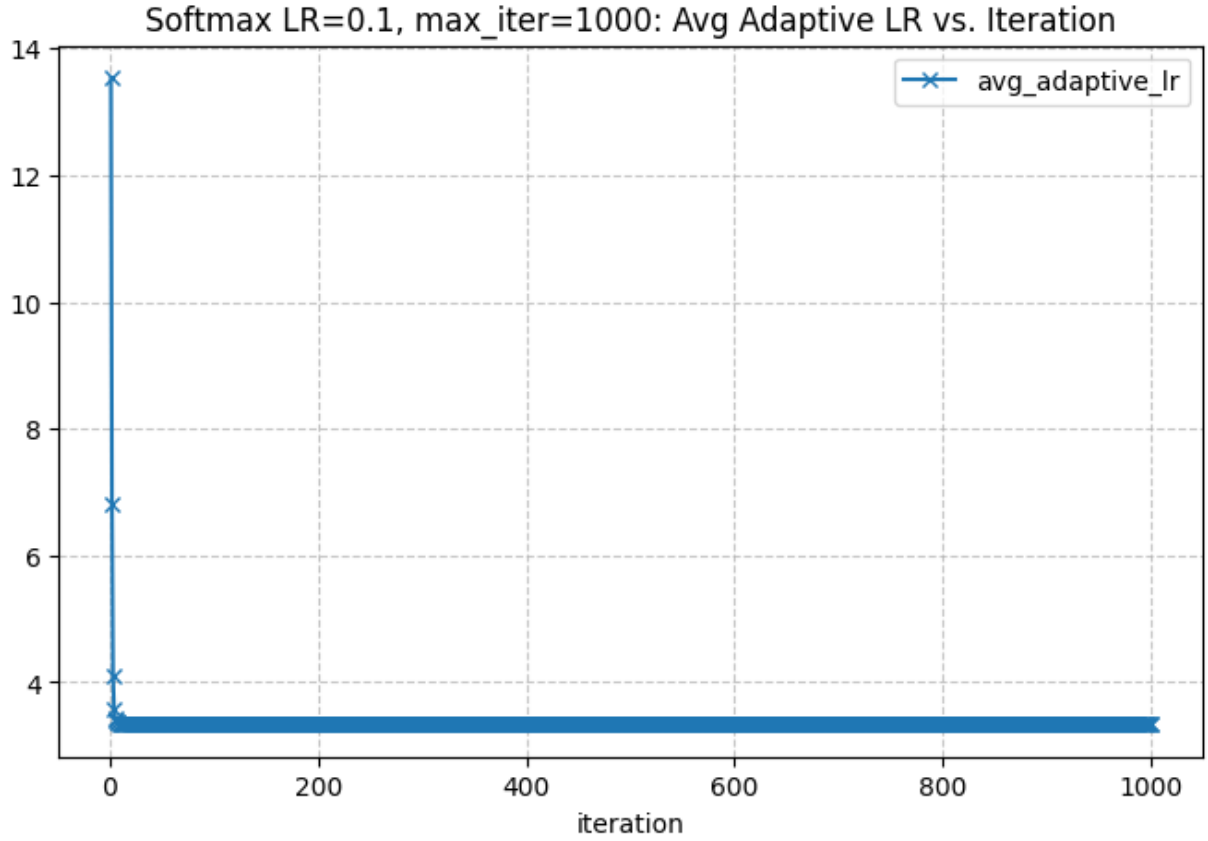

```
Confusions: Linear: 100%|██████████| 3/3 [00:00<00:00,  8.91it/s]
2025-03-18 21:51:55,880 - INFO - === Iteration-Level Visualization (All Models) ===
```

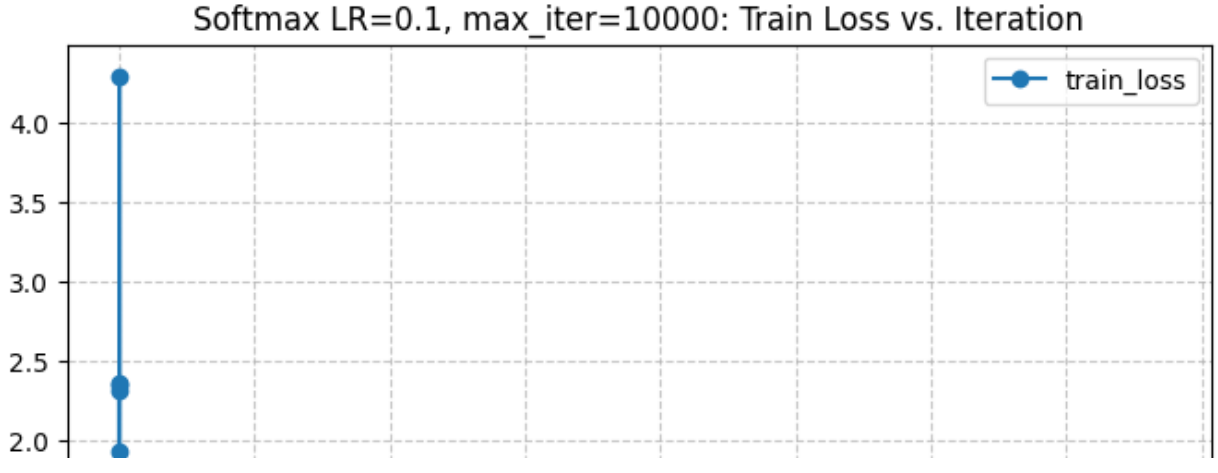


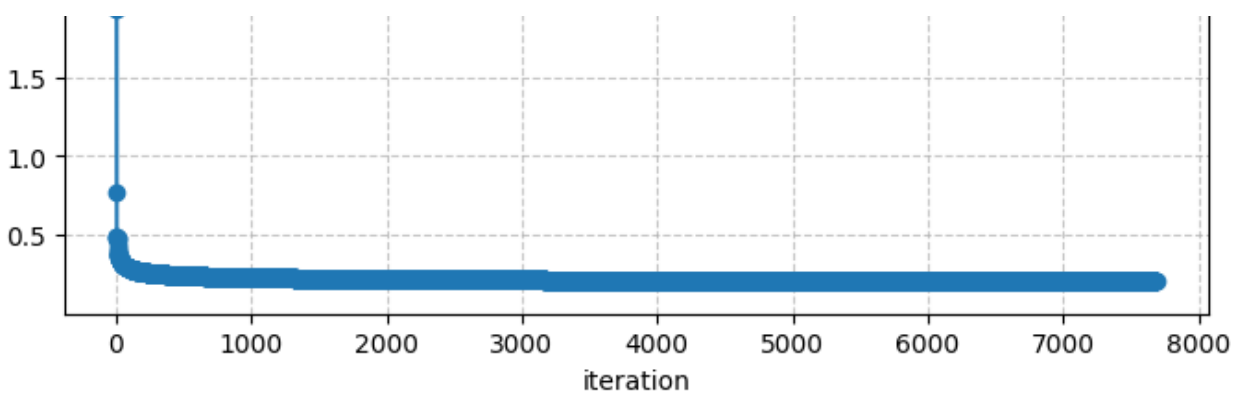



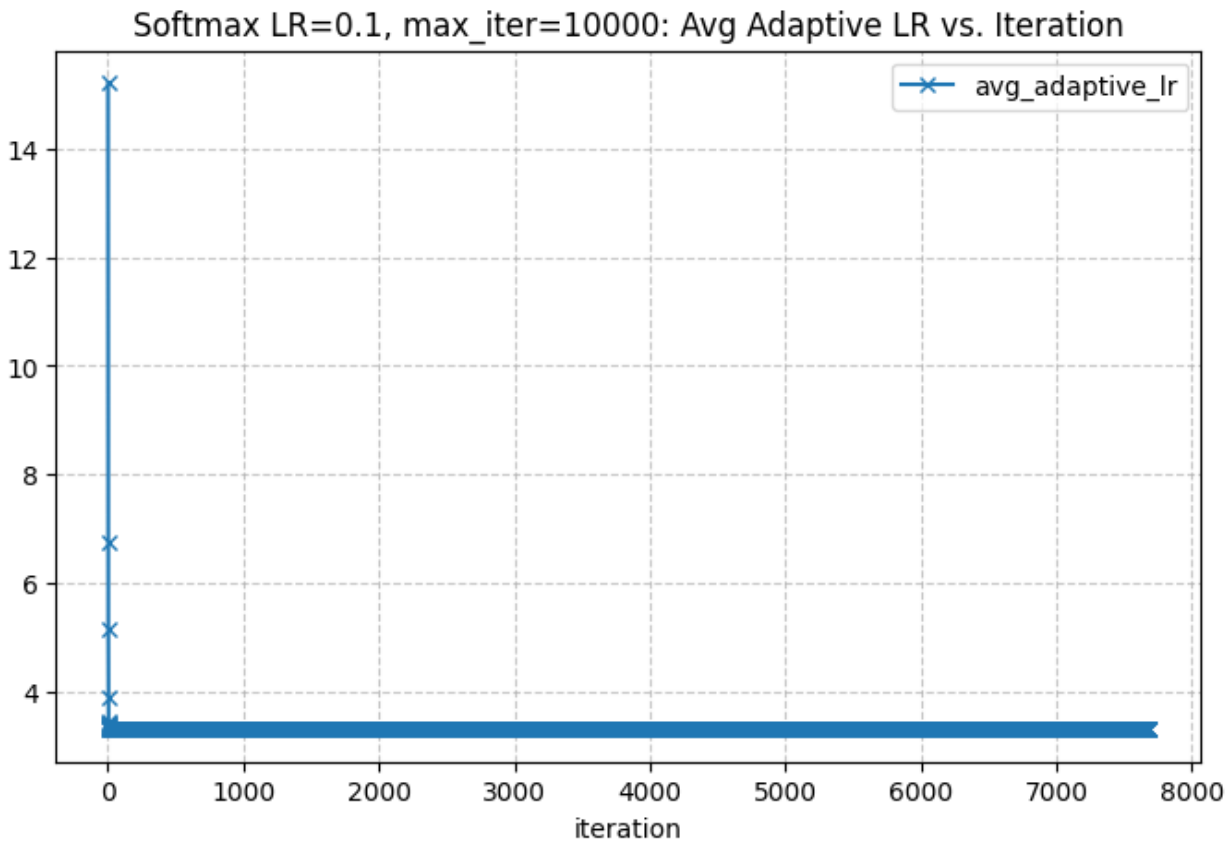

Softmax LR=0.1, max_iter=1000: Train Loss vs. Iteration

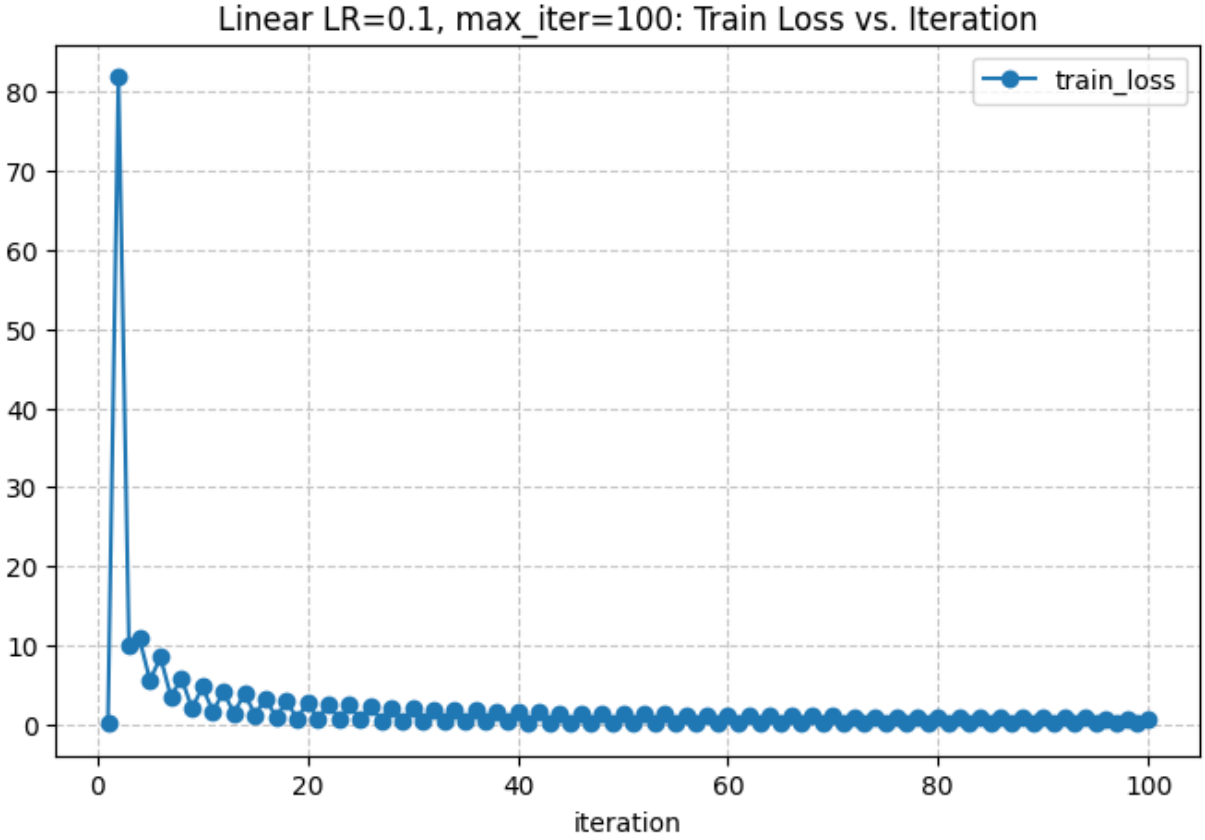

train_loss
4
3
2
1
0
200
400
600
800
1000
iteration

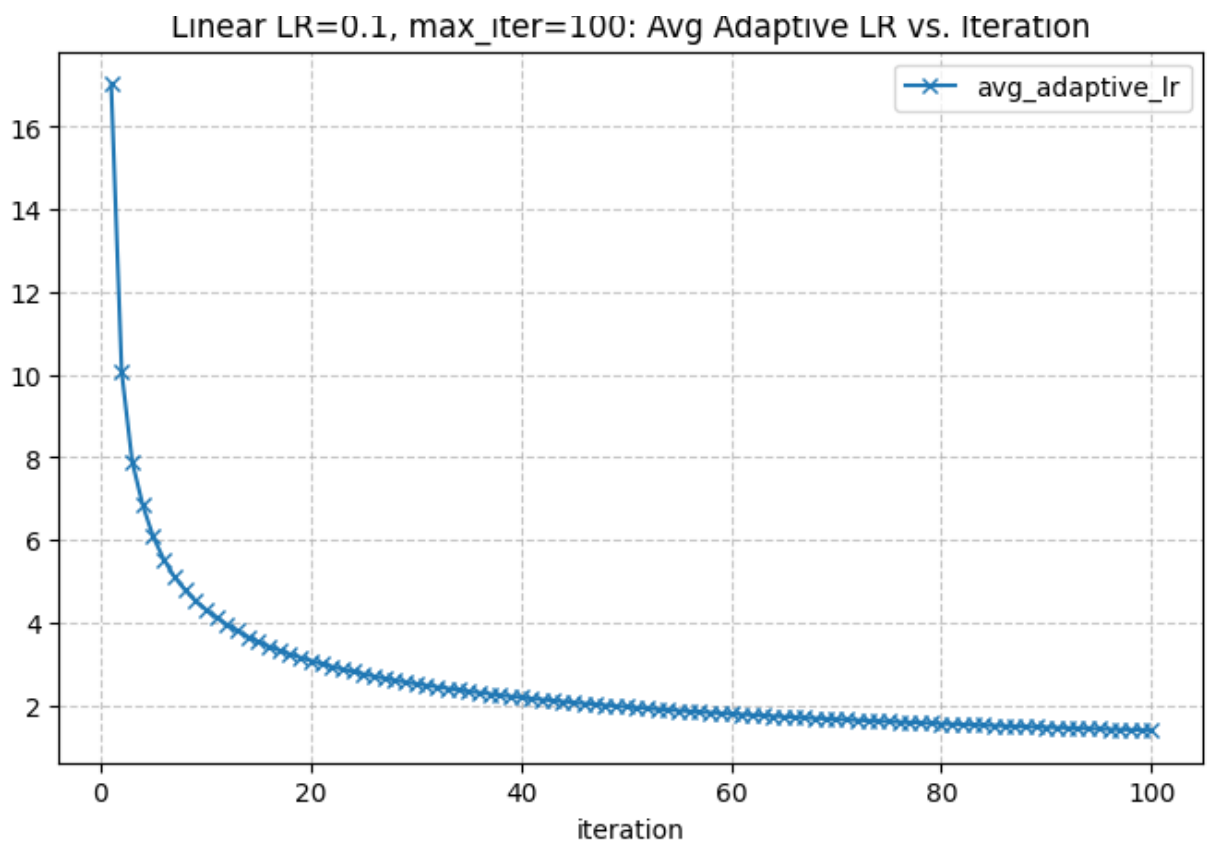

Softmax LR=0.1, max_iter=1000: Avg Adaptive LR vs. Iteration
avg_adaptive_lr
14
12
10
8
6
4
0
200
400
600
800
1000
iteration

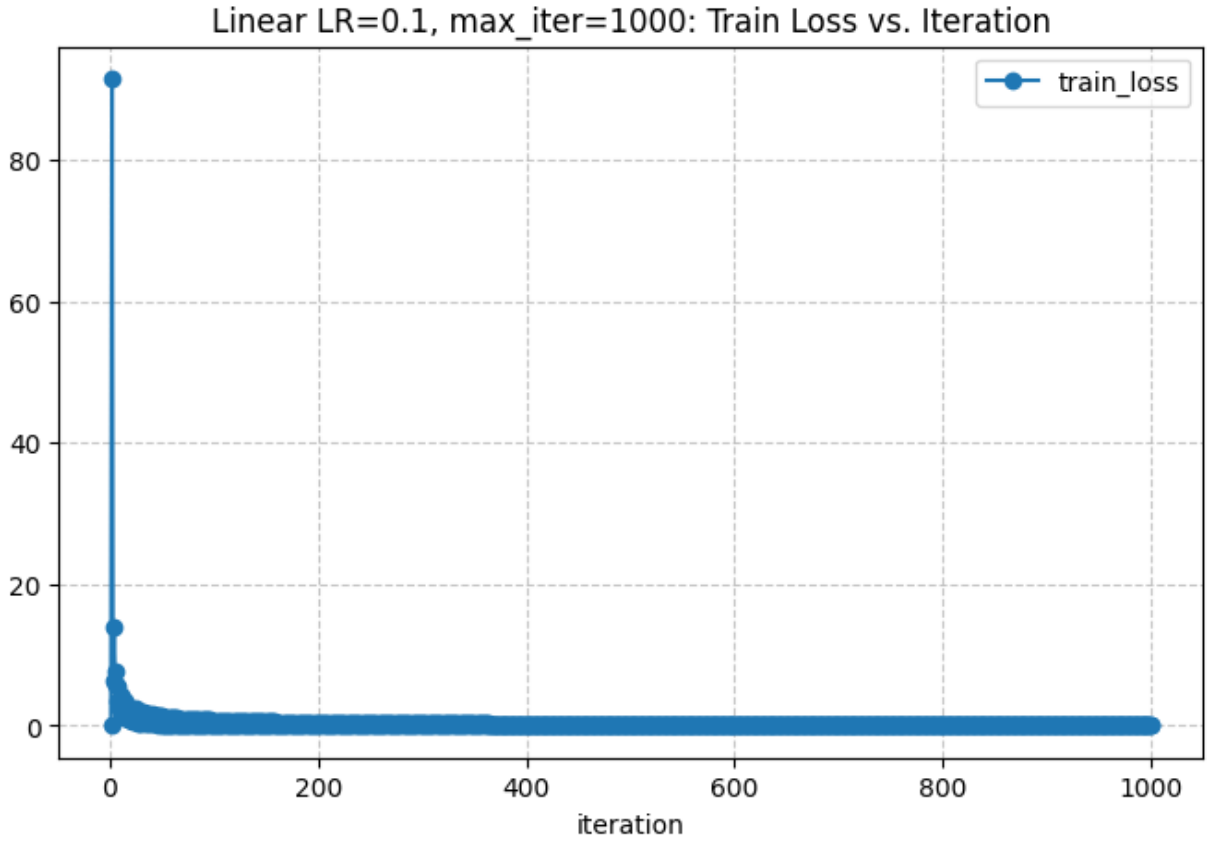

Softmax LR=0.1, max_iter=10000: Train Loss vs. Iteration
train_loss
4.0
3.5
3.0
2.5
2.0

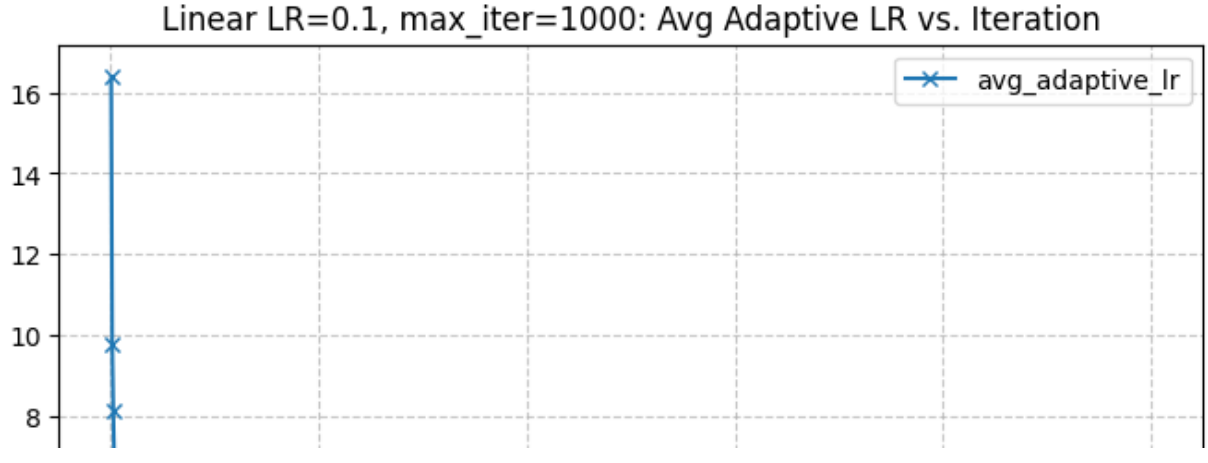

1.5
1.0
0.5
0
1000
2000
3000
4000
5000
6000
7000
8000
iteration

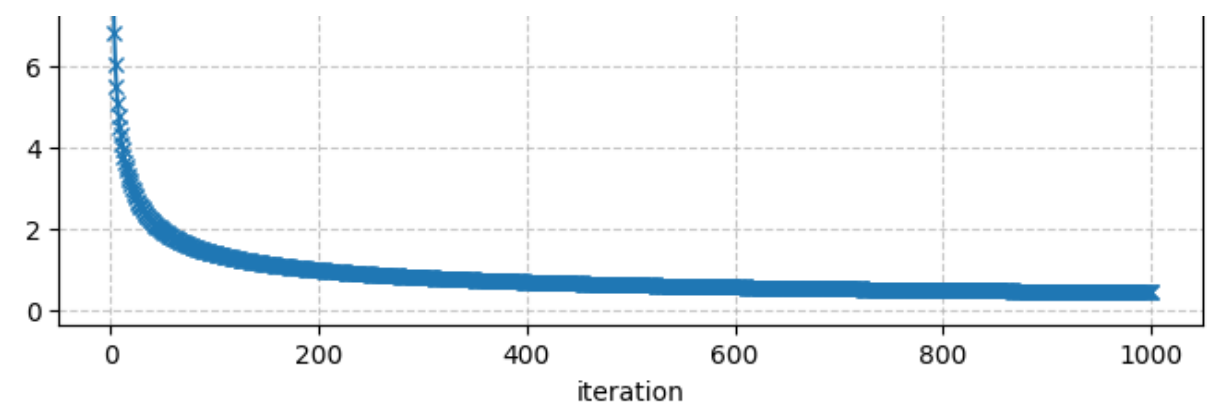

Softmax LR=0.1, max_iter=10000: Avg Adaptive LR vs. Iteration
avg_adaptive_lr
14
12
10
8
6
4
0
1000
2000
3000
4000
5000
6000
7000
8000
iteration

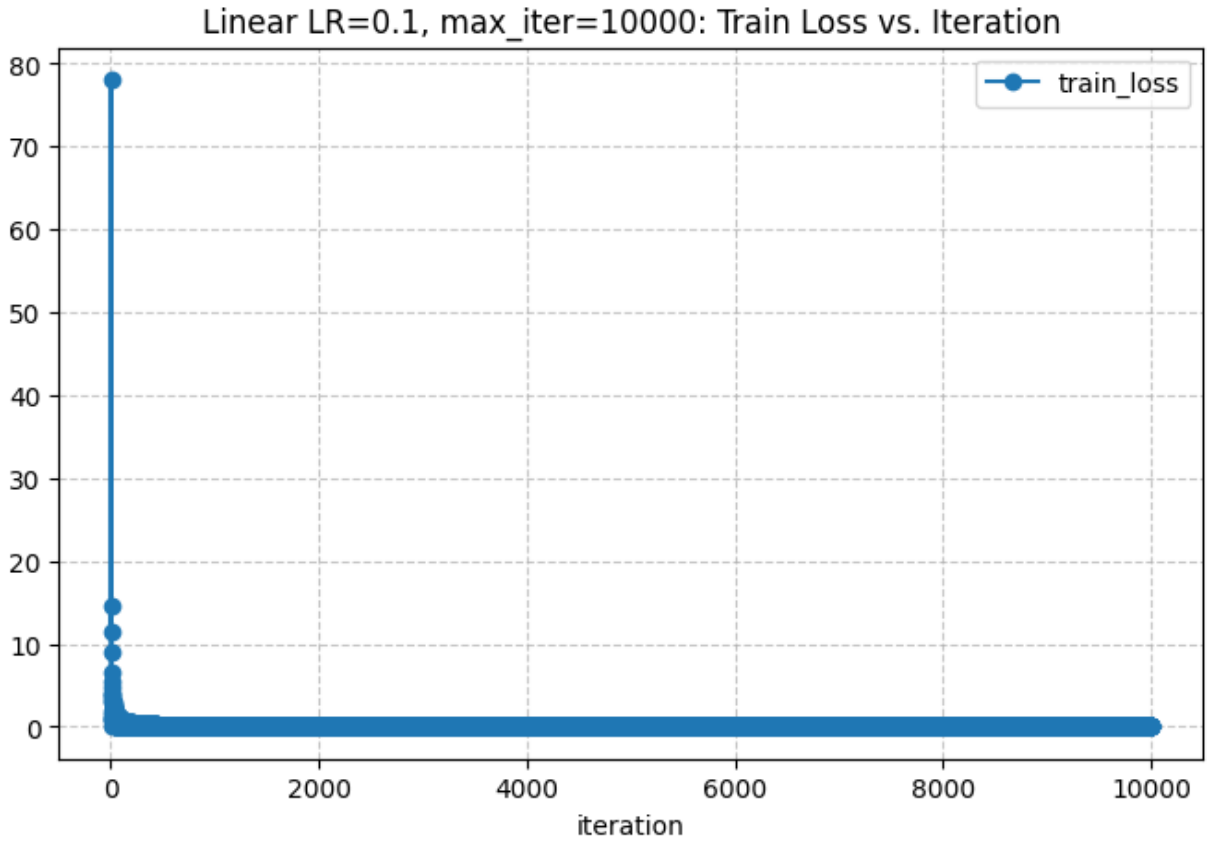

Linear LR=0.1, max_iter=100: Train Loss vs. Iteration
train_loss
80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
iteration

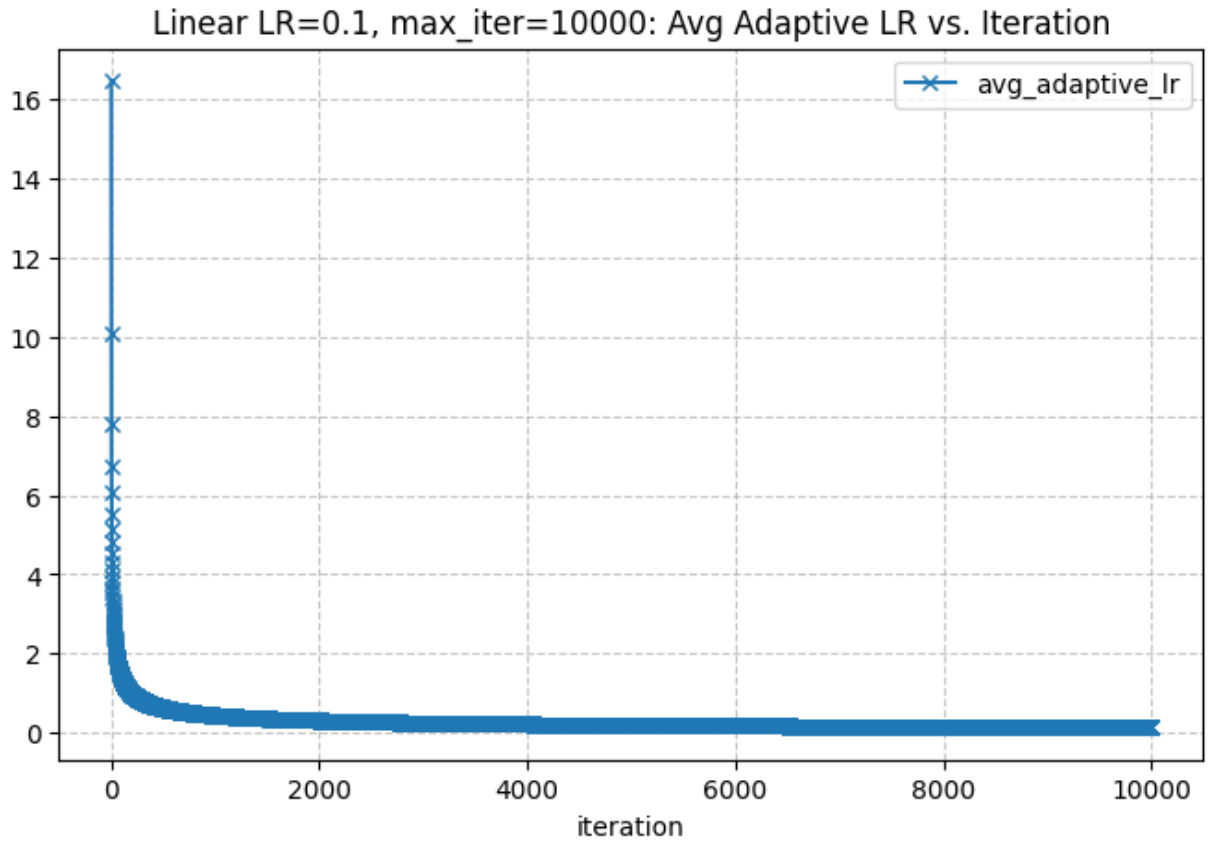

Linear LR=0.1, max_iter=100: Avg Adaptive LR vs. Iteration
avg_adaptive_lr
16
14
12
10
8
6
4
2
0
20
40
60
80
100
iteration


Linear LR=0.1, max_iter=1000: Train Loss vs. Iteration
train_loss
80
60
40
20
0
0
200
400
600
800
1000
iteration

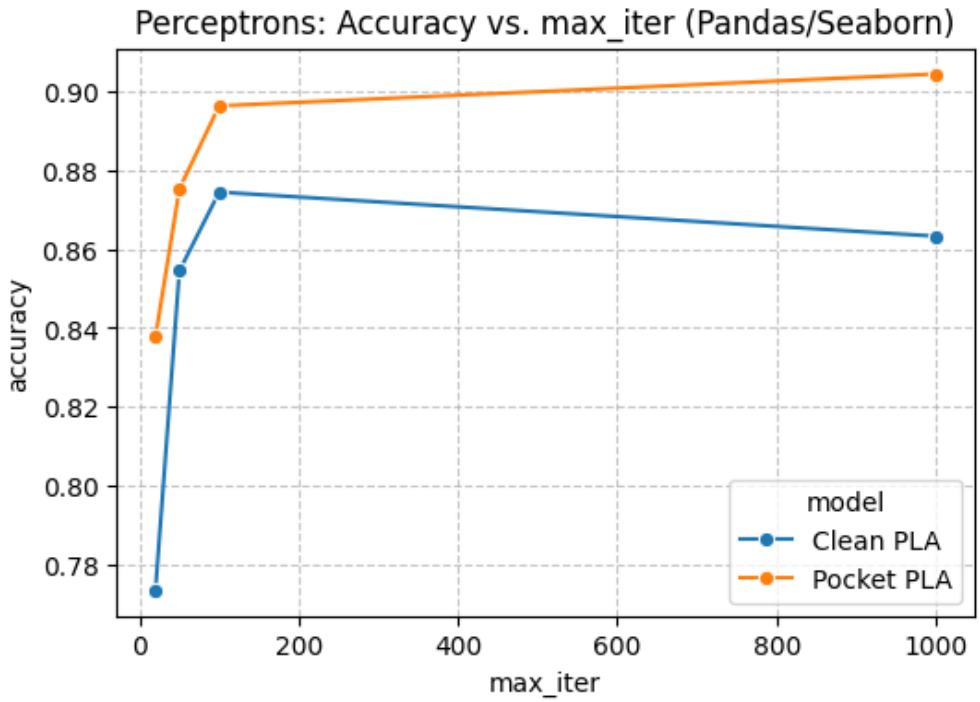

Linear LR=0.1, max_iter=1000: Avg Adaptive LR vs. Iteration
avg_adaptive_lr
16
14
12
10
8

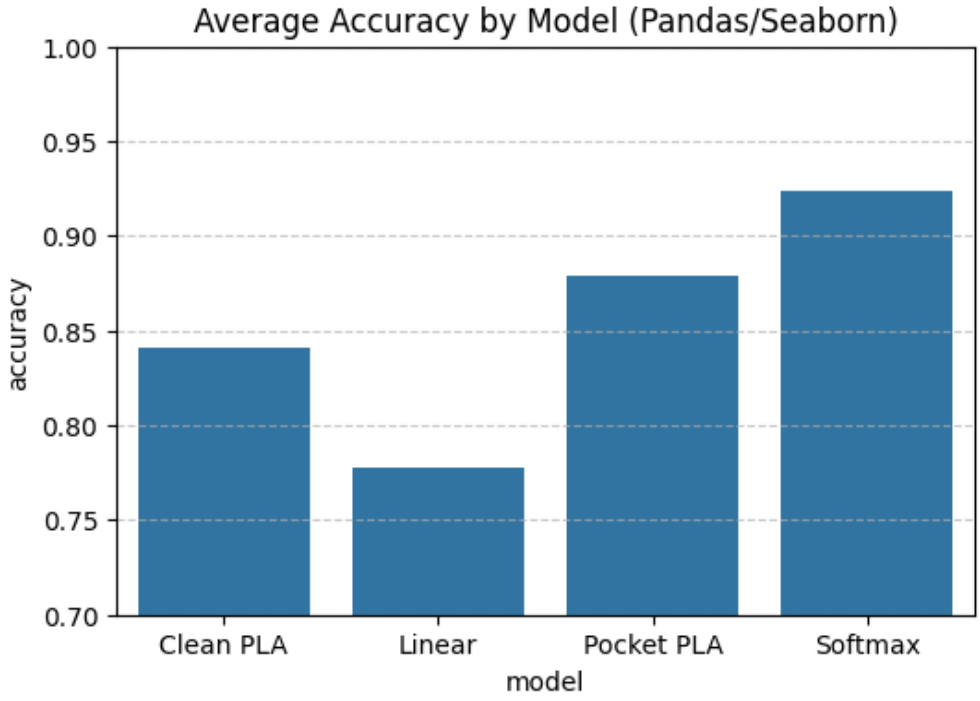

6
4
2
0
0
200
400
600
800
1000
iteration

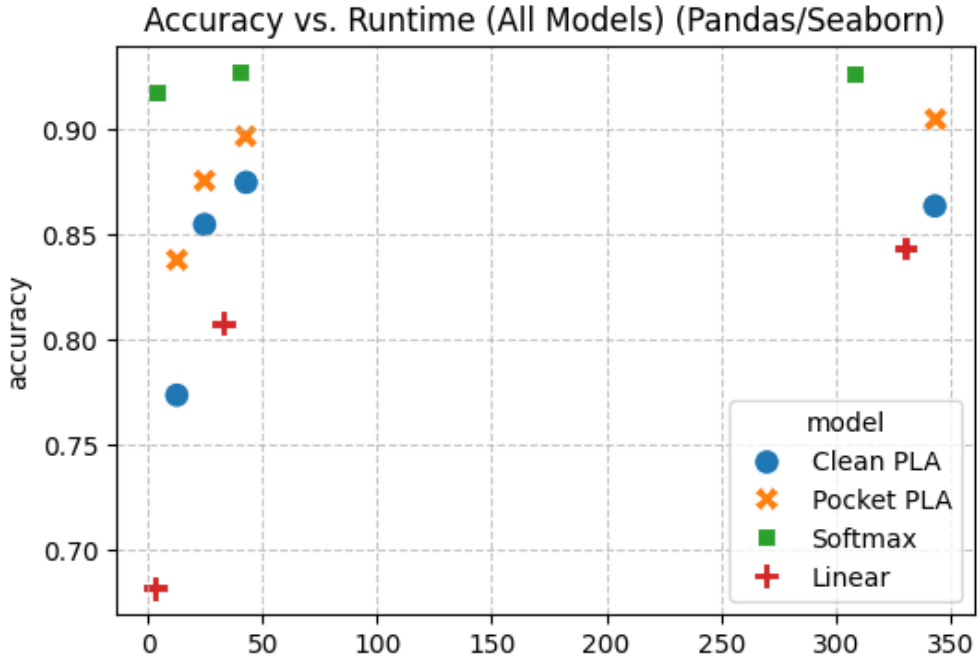

Linear LR=0.1, max_iter=10000: Train Loss vs. Iteration
train_loss
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
iteration


Linear LR=0.1, max_iter=10000: Avg Adaptive LR vs. Iteration
avg_adaptive_lr
16
14
12
10
8
6
4
2
0
0
2000
4000
6000
8000
10000
iteration

2025-03-18 21:51:56,609 - INFO - === Pandas + Seaborn Plots ===

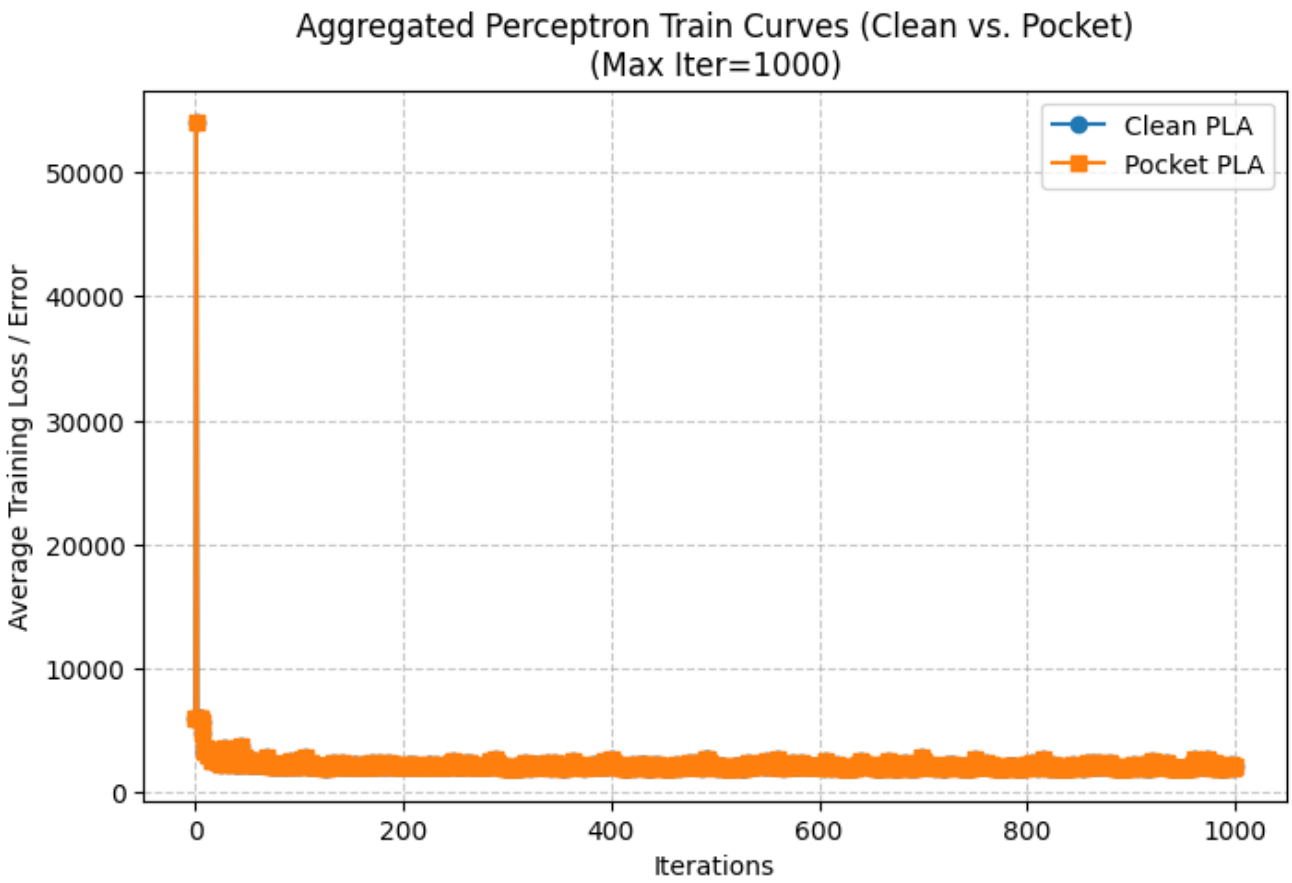


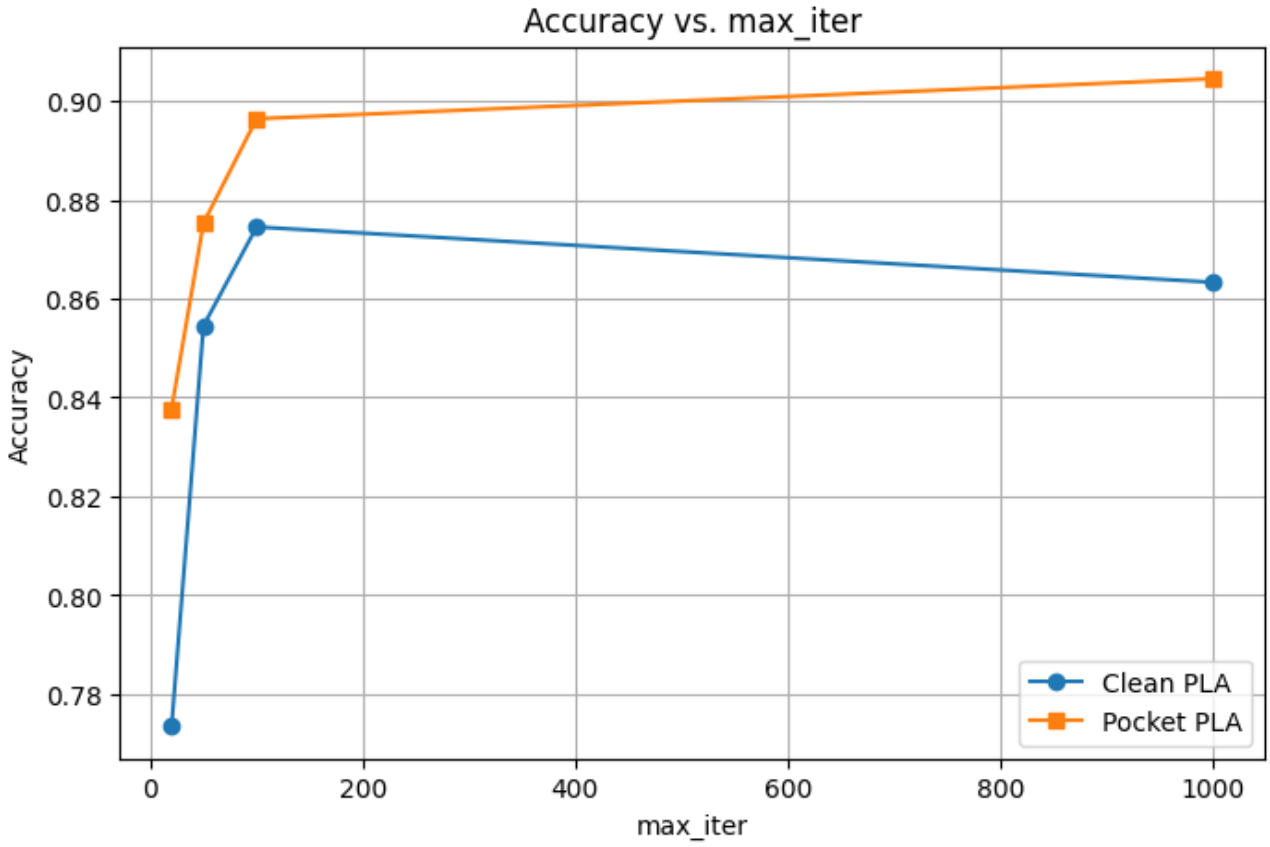



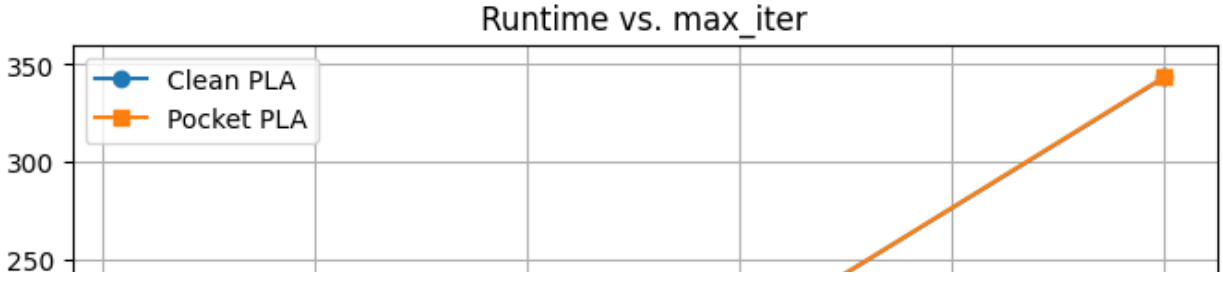



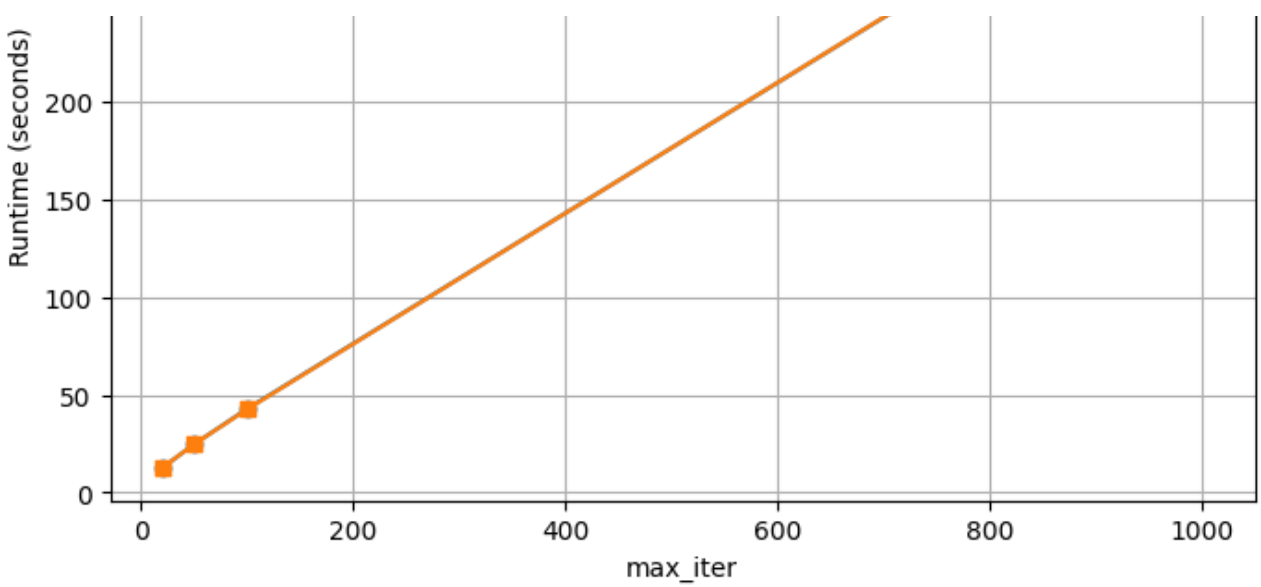
runtime

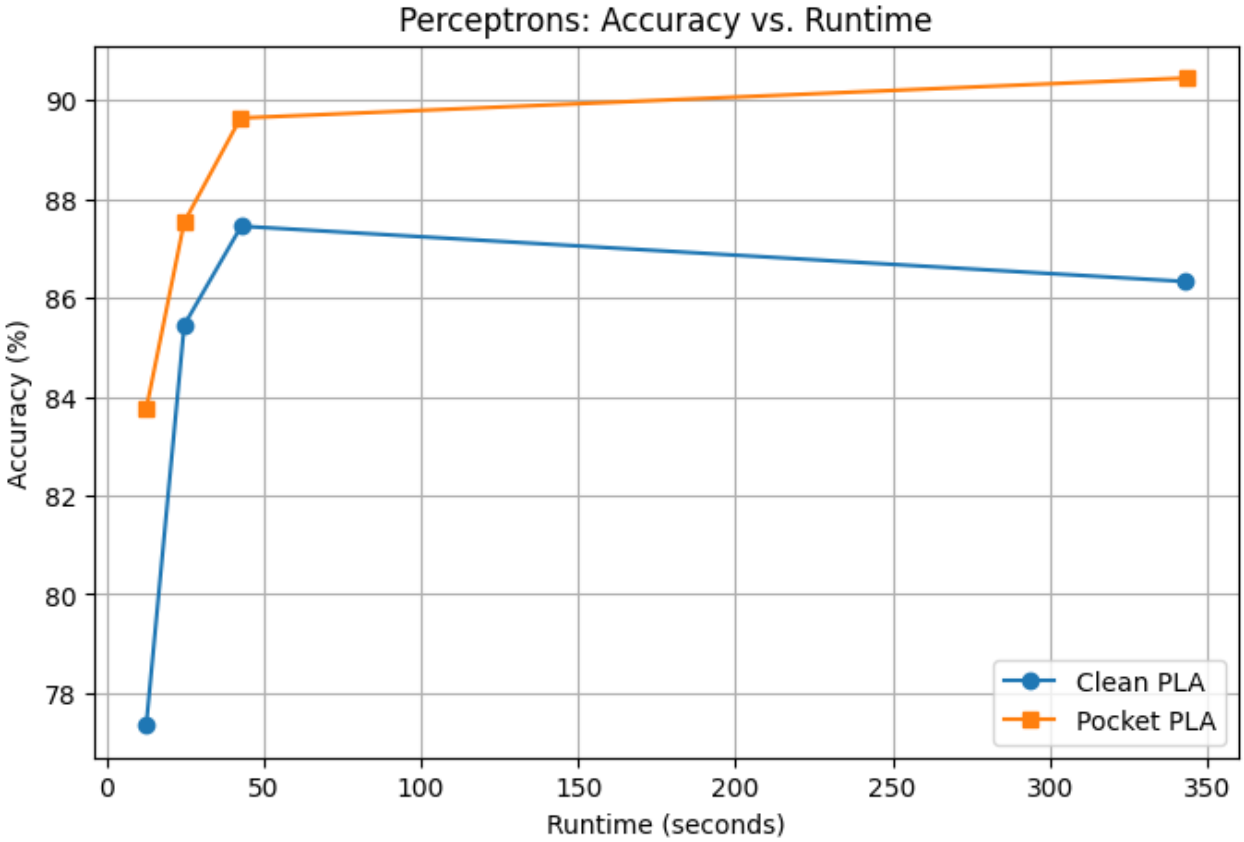
```
2025-03-18 21:51:56,806 - INFO - === Custom Summaries (Aggregated Curves, etc.) ===
```

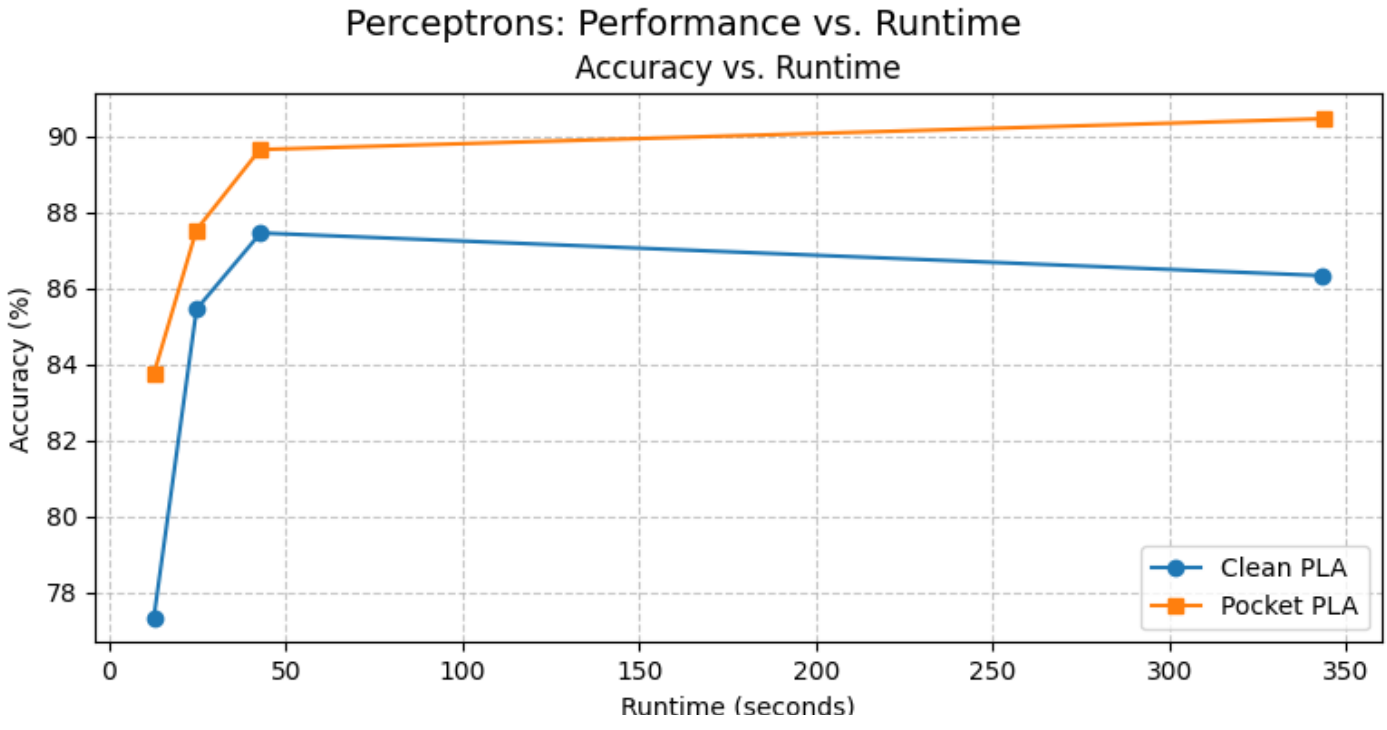


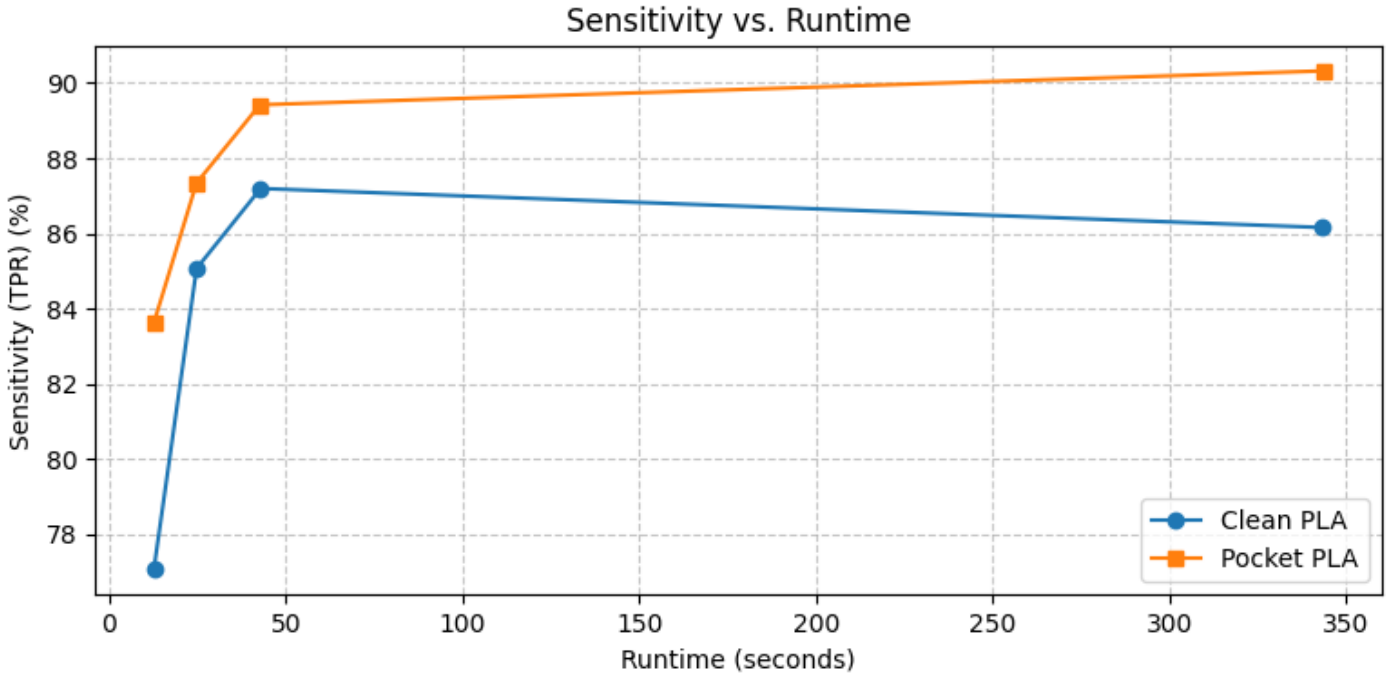



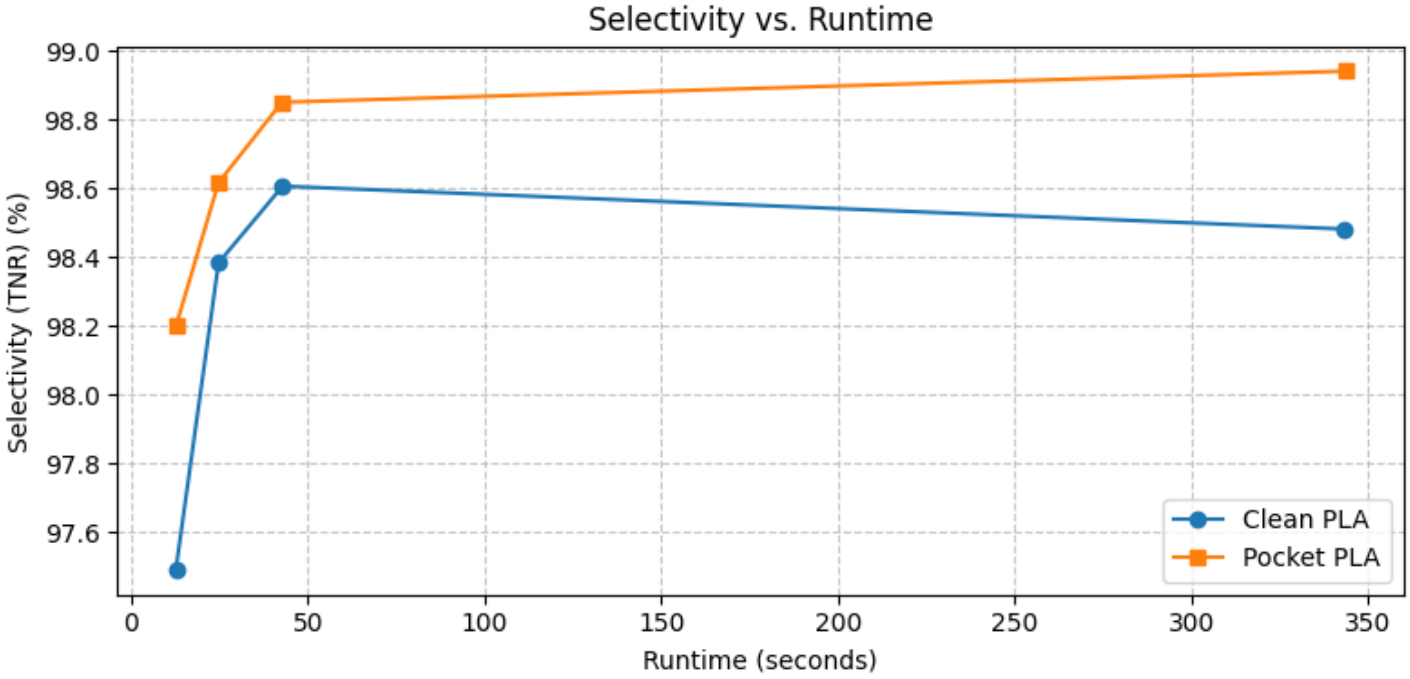



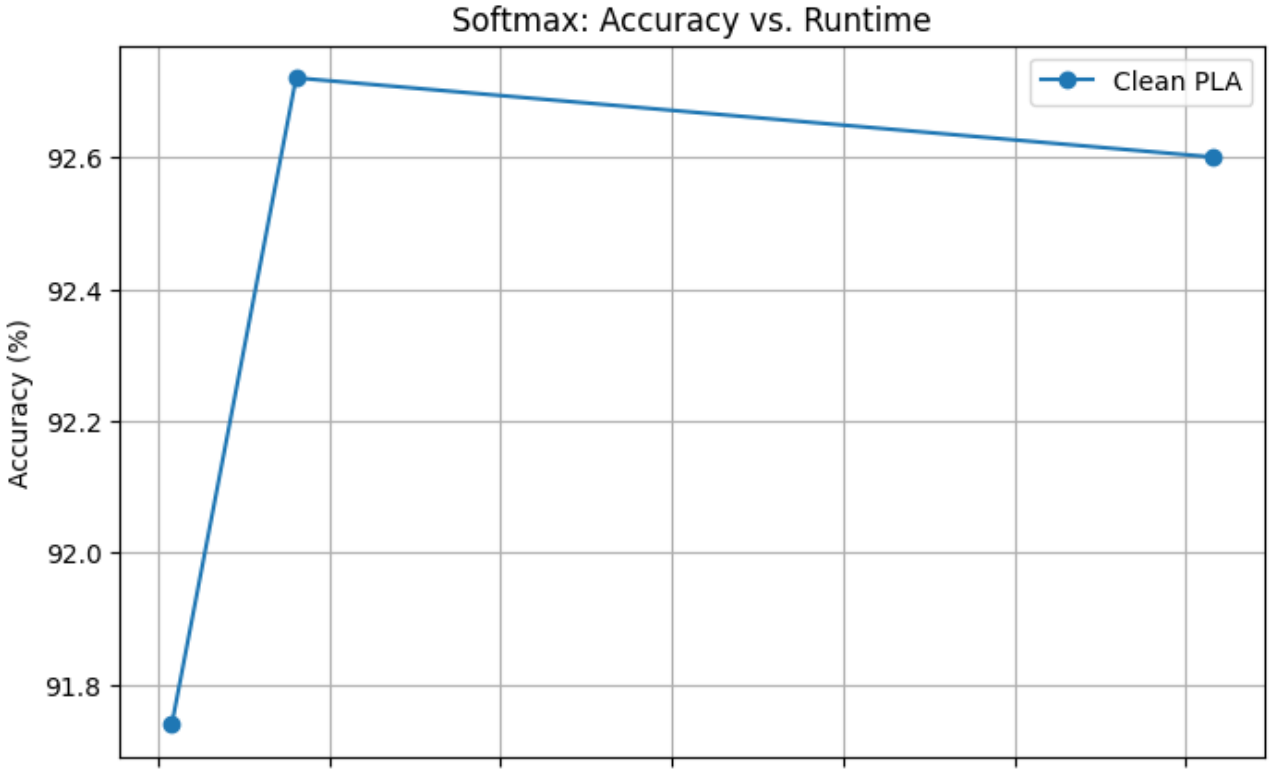

Runtime (seconds)
0
50
100
150
200
0
200
400
600
800
1000
max_iter

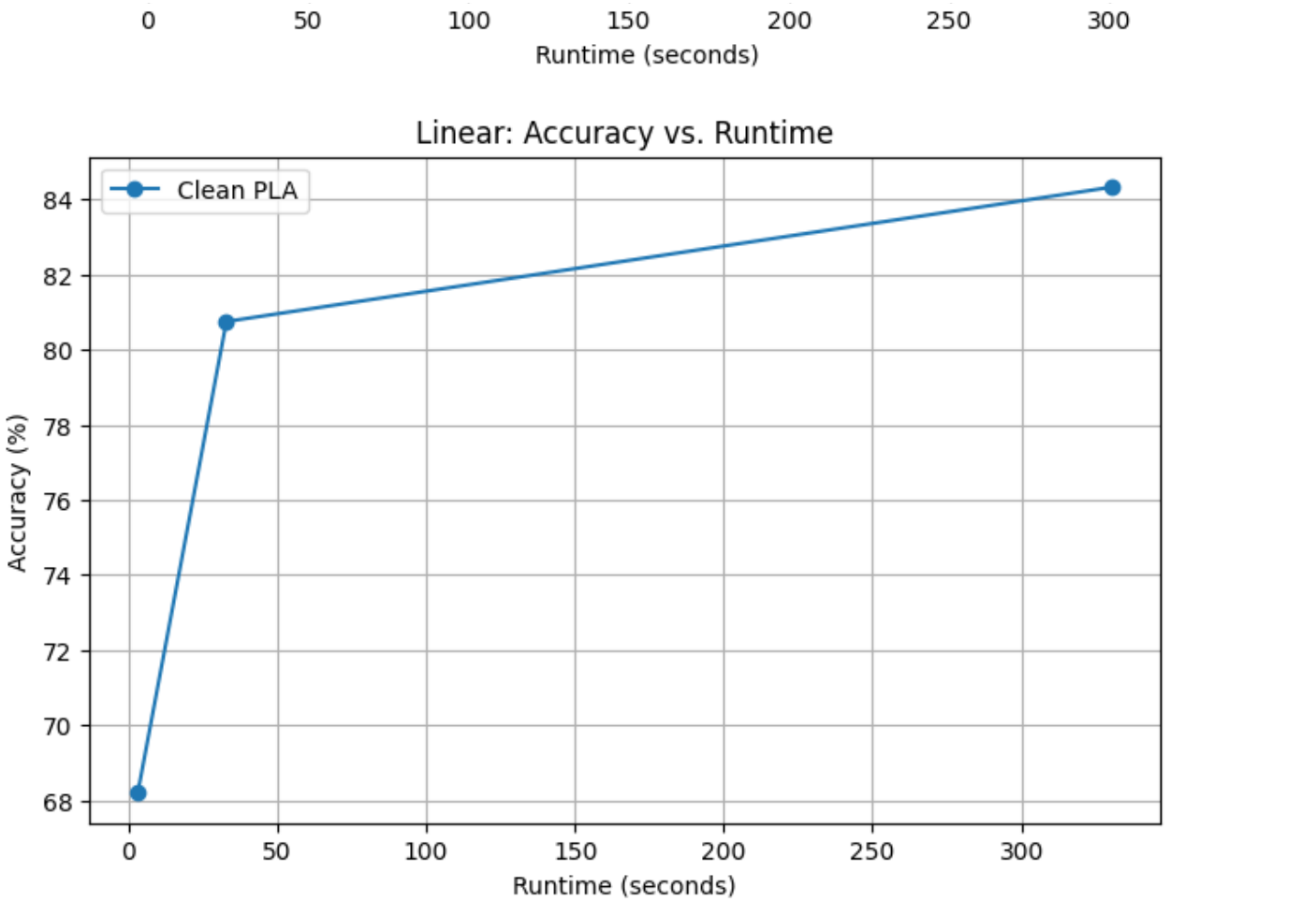

Perceptrons: Accuracy vs. Runtime
Accuracy (%)
78
80
82
84
86
88
90
0
50
100
150
200
250
300
350
Runtime (seconds)
Clean PLA
Pocket PLA

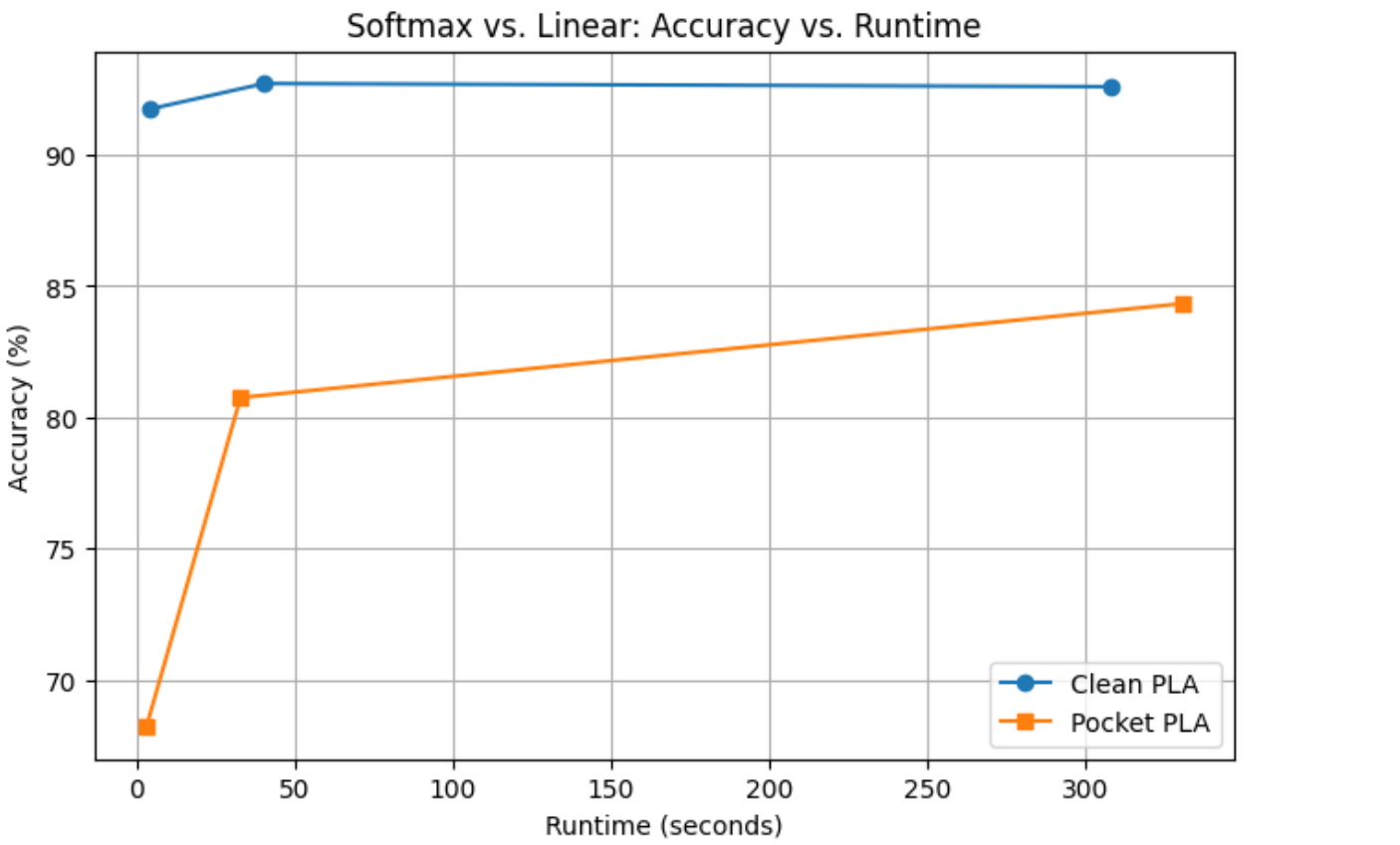

Perceptrons: Performance vs. Runtime
Accuracy vs. Runtime
Accuracy (%)
78
80
82
84
86
88
90
0
50
100
150
200
250
300
350
Runtime (seconds)
Clean PLA
Pocket PLA

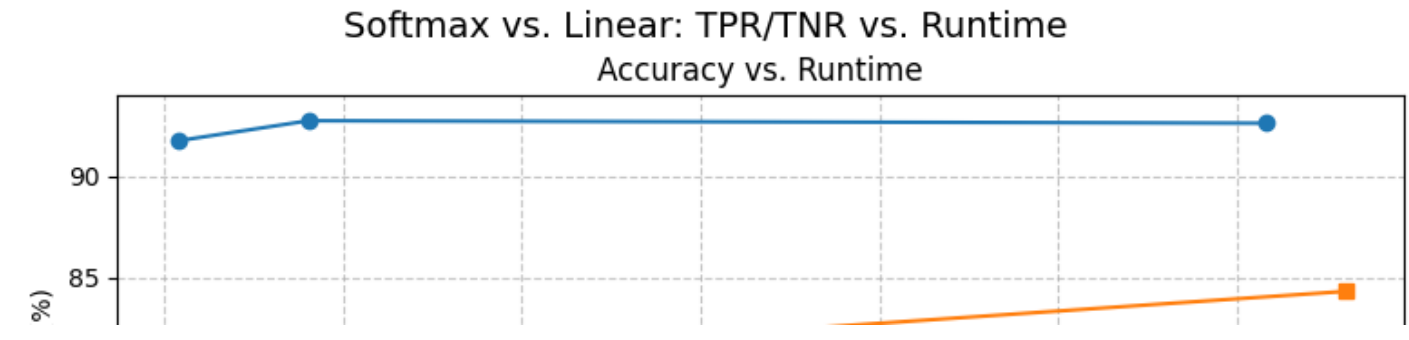

Sensitivity vs. Runtime
Sensitivity (TPR) (%)
Runtime (seconds)
Clean PLA
Pocket PLA

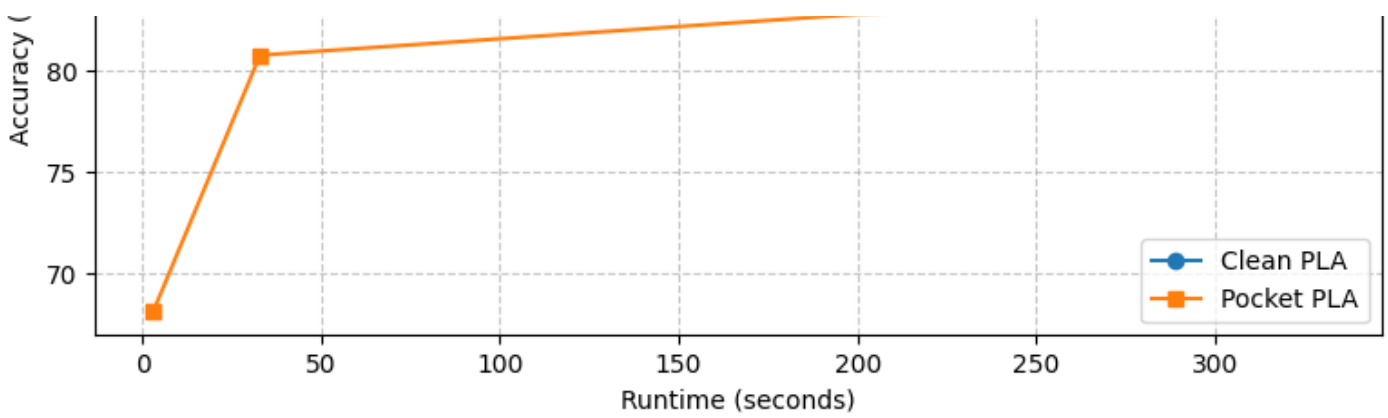

Selectivity vs. Runtime
Selectivity (TNR) (%)
Runtime (seconds)
Clean PLA
Pocket PLA

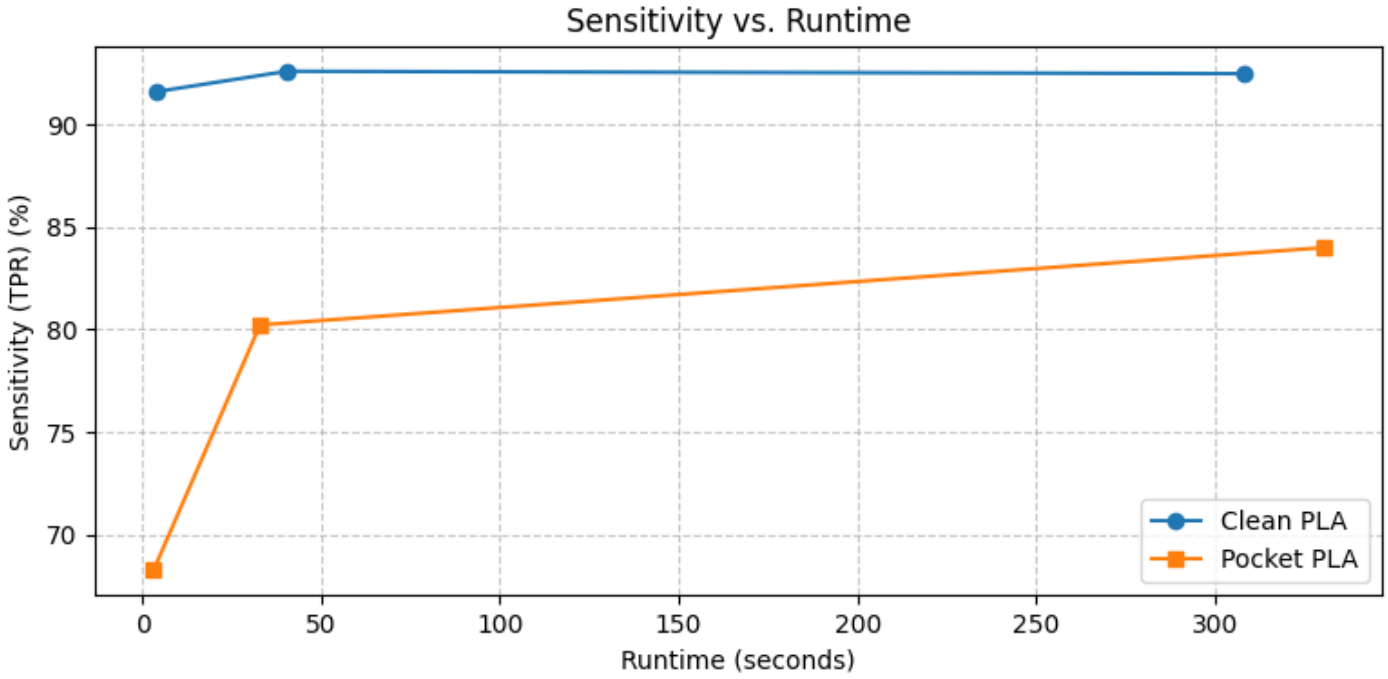

Softmax: Accuracy vs. Runtime
Accuracy (%)
Clean PLA

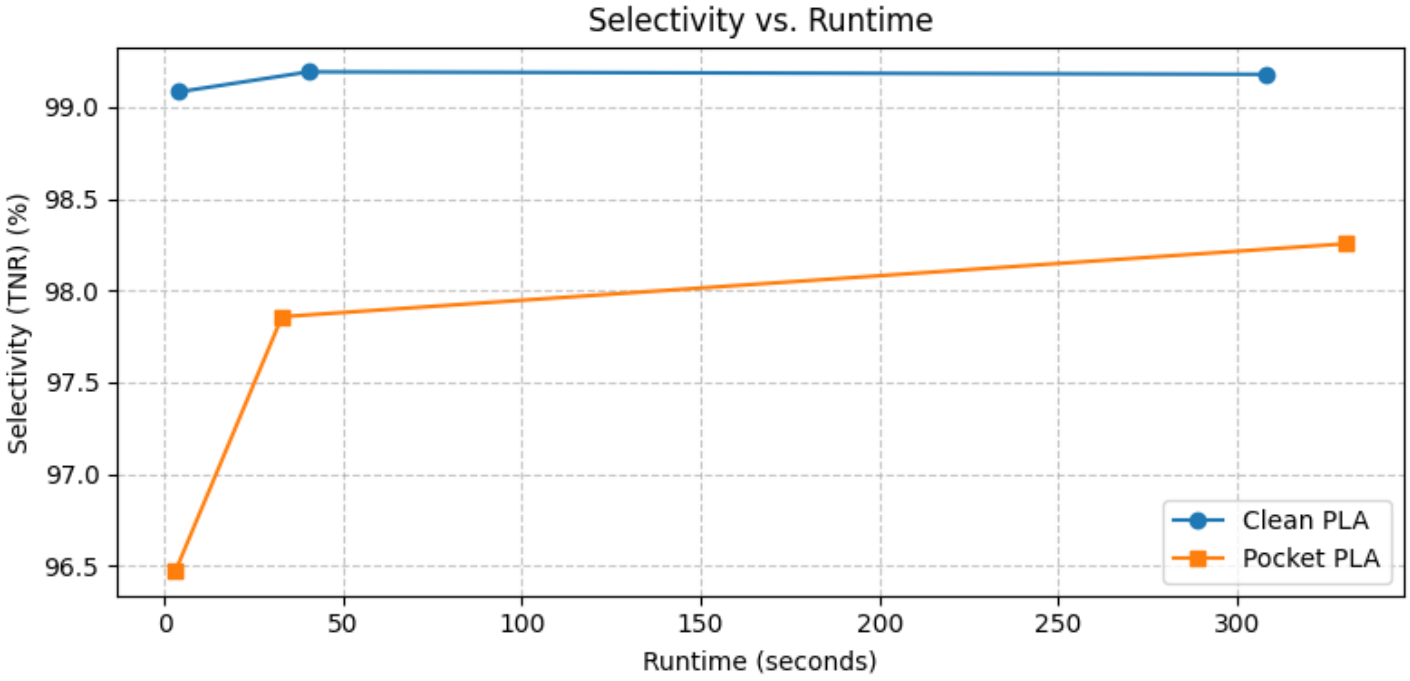

0
50
100
150
200
250
300
Runtime (seconds)
Linear: Accuracy vs. Runtime
Clean PLA
84
82
80
78
76
74
72
70
68
Accuracy (%)
0
50
100
150
200
250
300
Runtime (seconds)

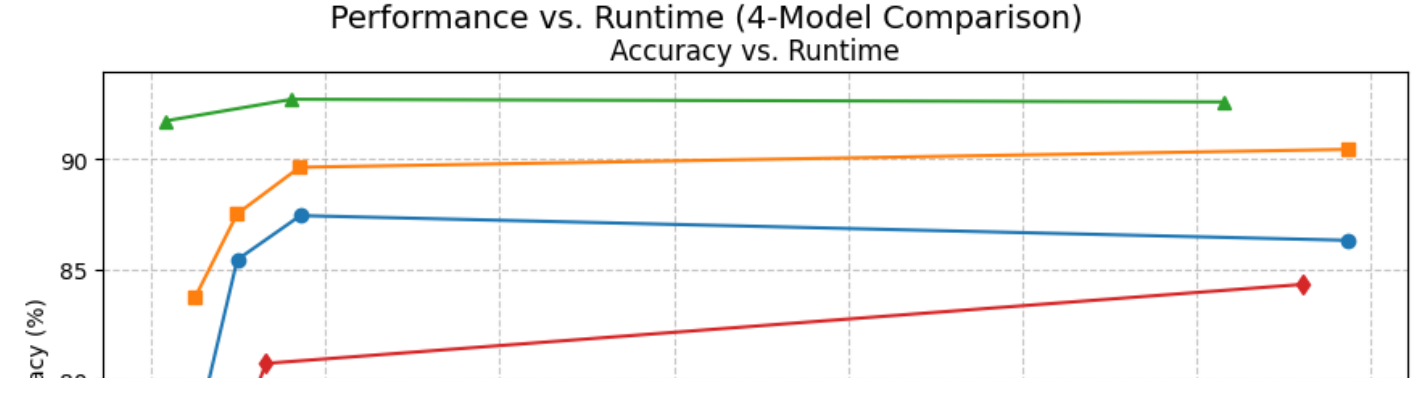

Softmax vs. Linear: Accuracy vs. Runtime
90
85
80
75
70
Accuracy (%)
Clean PLA
Pocket PLA
0
50
100
150
200
250
300
Runtime (seconds)

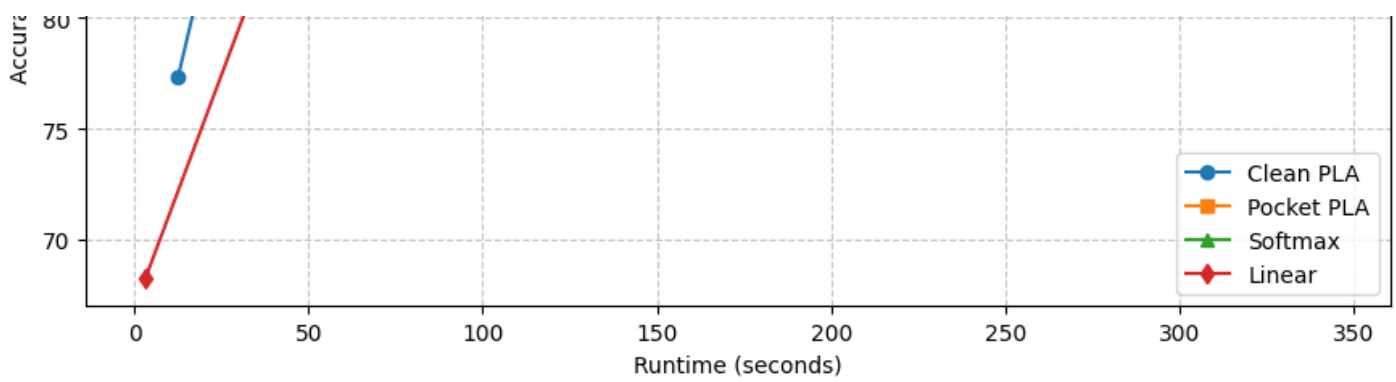

Softmax vs. Linear: TPR/TNR vs. Runtime
Accuracy vs. Runtime
90
85

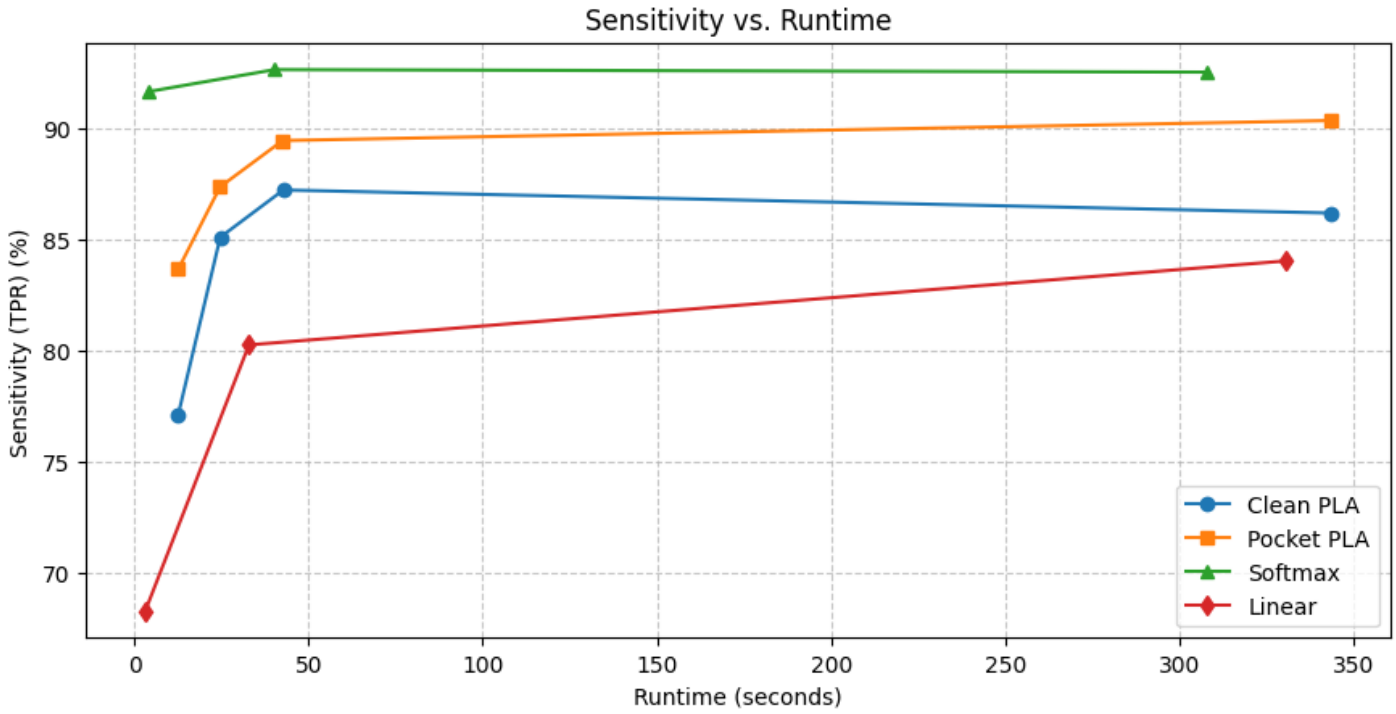

Accuracy (
80
75
70
0
50
100
150
200
250
300
Runtime (seconds)
Clean PLA
Pocket PLA

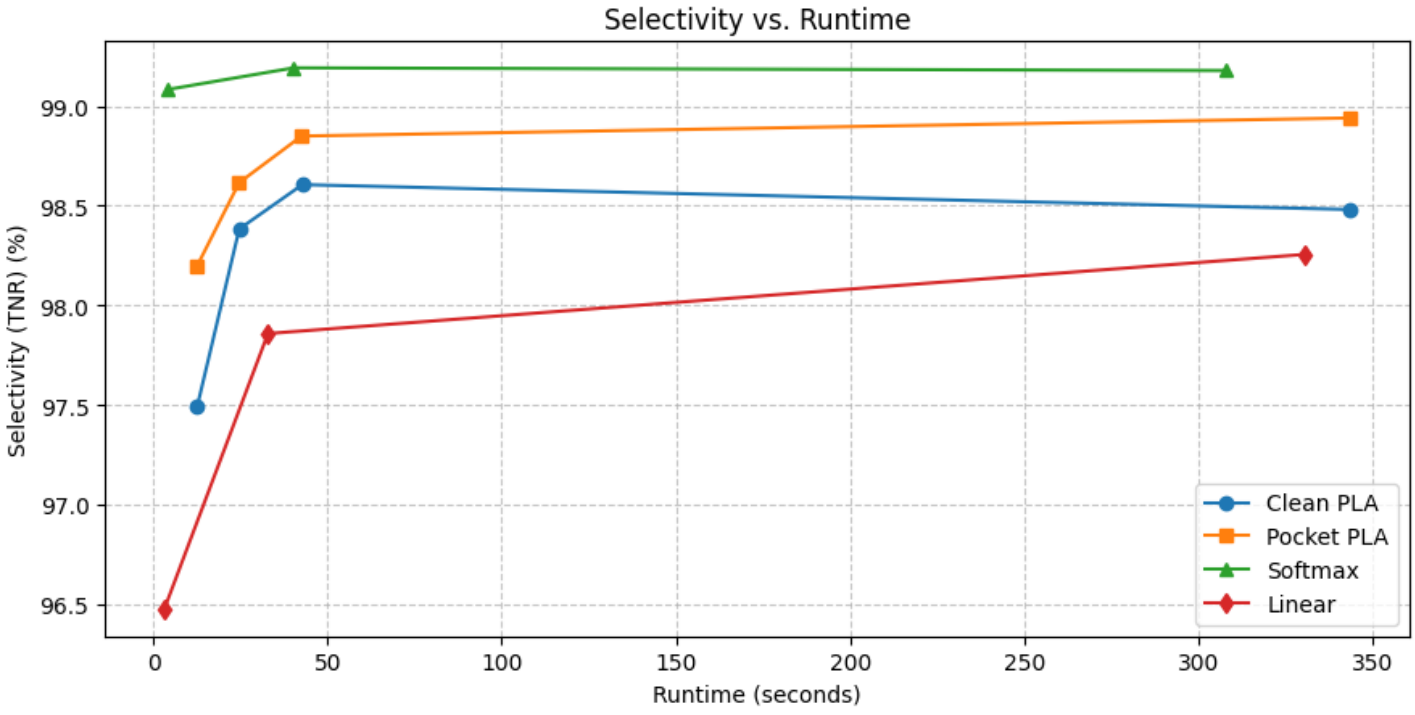

Sensitivity vs. Runtime
Sensitivity (TPR) (%)
90
85
80
75
70
0
50
100
150
200
250
300
Runtime (seconds)
Clean PLA
Pocket PLA

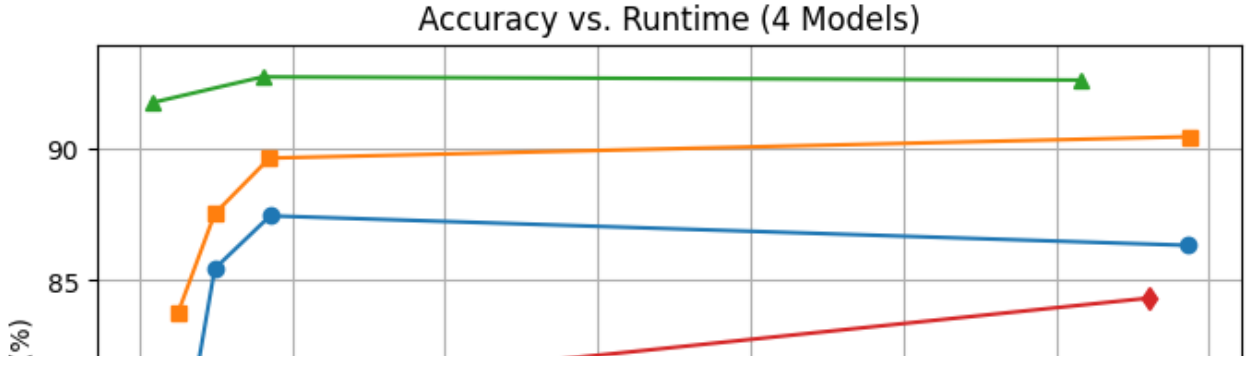

Selectivity vs. Runtime
Selectivity (TNR) (%)
99.0
98.5
98.0
97.5
97.0
96.5
0
50
100
150
200
250
300
Runtime (seconds)
Clean PLA
Pocket PLA

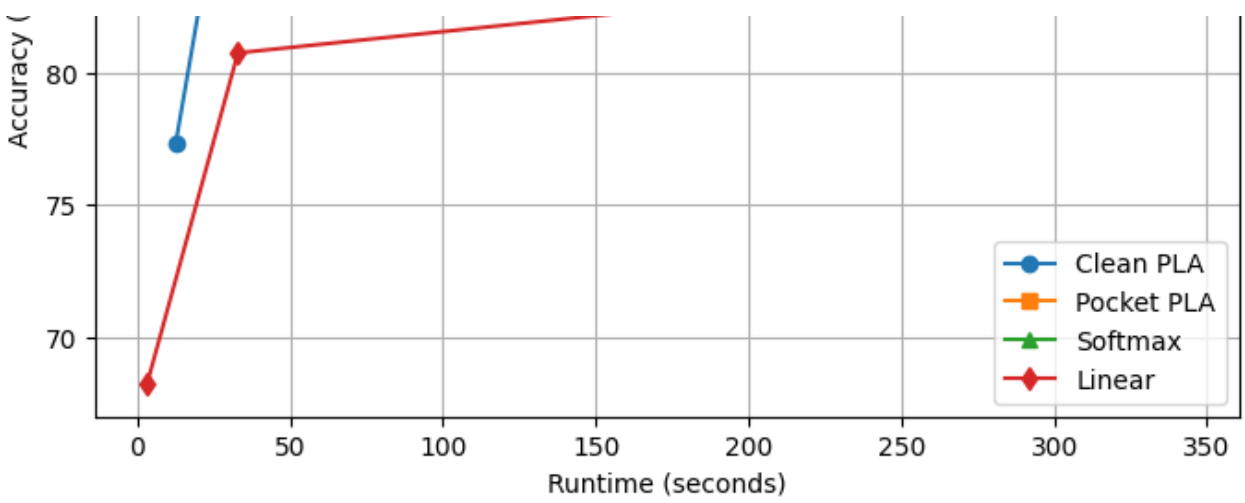

Performance vs. Runtime (4-Model Comparison)
Accuracy vs. Runtime
90
85
acy (%)


Accura
80
75
70
0
50
100
150
200
250
300
350
Runtime (seconds)
Clean PLA
Pocket PLA
Softmax
Linear


Sensitivity vs. Runtime
Sensitivity (TPR) (%)
90
85
80
75
70
0
50
100
150
200
250
300
350
Runtime (seconds)
Clean PLA
Pocket PLA
Softmax
Linear


Selectivity vs. Runtime
Selectivity (TNR) (%)
99.0
98.5
98.0
97.5
97.0
96.5
0
50
100
150
200
250
300
350
Runtime (seconds)
Clean PLA
Pocket PLA
Softmax
Linear


Accuracy vs. Runtime (4 Models)
90
85




```
2025-03-18 21:51:57,814 - INFO - === All Visualizations Complete ===
```

# Final Results Summary

**Observations:**

- **Pocket PLA** consistently outperforms Clean PLA in both accuracy and sensitivity (TPR) across all tested iteration counts.
- Increasing `max_iter` improves performance, though gains tend to plateau beyond roughly 50–100 iterations.
- **Runtime** increases nearly linearly with `max_iter` for both methods, highlighting a clear trade-off between higher accuracy and computational cost.
- Perfect linear separation is not achieved—even at higher iteration counts, neither method reaches 100% accuracy, indicating that the dataset is not strictly linearly separable.

**Trade-off Analysis:**

- **Low Iterations (max_iter = 10–30):**
  Fast training with modest accuracy and TPR, suitable for rapid prototyping or time-sensitive applications.
- **Medium Iterations (max_iter = 50–100):**
  Balanced performance and runtime, capturing most achievable gains without excessive overhead.
- **High Iterations (max_iter > 100):**
  Marginal performance improvements with significant runtime increase; diminishing returns for practical applications.

**Recommendations for Future Work:**

- Experiment with alternative update rules (e.g., adaptive learning rates) to accelerate convergence.
- Compare against more sophisticated models (e.g., Logistic Regression, SVMs, neural networks) for broader insights.
- Evaluate model robustness under noisy or adversarial conditions.

This comprehensive analysis—including confusion matrices, error curves, and summary plots—provides detailed insights into the performance of the multi-class Perceptron on MNIST and informs the optimal balance between training efficiency and classification performance.