Overview

- Agents must be capable of learning and using information immediately.
- Deep neural networks (DNNs) are widely used for perception tasks, but if they are updated on changing data distributions, they catastrophically forget previous knowledge.
- Streaming learning requires agents to learn from non-independent and identically distributed (iid) data streams in real-time, i.e., one example at a time and a single pass through the dataset.
- Deep Streaming Linear Discriminant Analysis (SLDA) trains the output layer of a convolutional neural network (CNN) incrementally.
- SLDA outperforms recent incremental batch and streaming models with fewer memory and computational costs.

Deep Streaming Linear Discriminant Analysis

- SLDA stores a running mean per class \( \mu_k \) and a tied covariance matrix \( C \).
- We compute the precision matrix \( \Lambda = \left((1 - e^2)\Sigma + eI\right)^{-1} \).
- Predictions are made by assigning to an input embedding \( z_t \) the label of the closest Gaussian in feature space using the stored means and covariance:
  \[
  \hat{y}_t = \arg\max_k \left( \mu_k^T z_t - \frac{1}{2} (z_t - \mu_k)^T \Lambda_{kk} (z_t - \mu_k) \right).
  \]

Experimental Evaluation

- **ImageNet-1K**: Popular large-scale image classification dataset (1,000 classes).
- **CORE50**: Streaming dataset containing video sequences of 10 different object categories. Temporal dependences are natural for streaming.

\[
\rho_{all} = \frac{1}{T} \sum_{t=1}^{T} \frac{a_t}{a_{t, offline}}, \quad a_t = \text{accuracy of streaming learner at time } t \quad a_{t, offline} = \text{accuracy of offline model at time } t.
\]

Summary

- SLDA is popular in the data mining community but has not been used recently for large classification datasets.
- We combine SLDA with a CNN and exceed incremental batch learning models, while being much more lightweight.
- Our offline results suggest greater performance is achievable by training hidden layers, but we urge future developers to test only training the output layer to ensure gains are being realized.

Acknowledgements

We thank DARPA L2M for financially supporting this research.