

Confronto tra Adam, Adagrad e SVRG su classificazione Cats vs Dogs

Fabio Bozzoli

Unimore

7 settembre 2025

Obiettivi dell'esperimento

- Confrontare tre algoritmi di ottimizzazione:
 - Adam
 - Adagrad
 - SVRG (Stochastic Variance Reduced Gradient)
- Valutare:
 - Accuratezza di classificazione
 - Stabilità del training
 - Tempo di addestramento
- Discutere il significato delle scelte progettuali

Setup sperimentale

- **Dataset:** Sottoinsieme Cats vs Dogs (2000 immagini train, 500 test)
- **Modello:** Rete fully connected
 - Input: 128x128x3
 - Hidden: 500 neuroni, ReLU
 - Output: 1 neurone (classificazione binaria)
- **Loss:** Hinge loss (tipo SVM)
- **Iperparametri:**
 - Learning rate: {0.005, 0.001}
 - Batch size: {64, 128}
 - Epoche: 10

Algoritmi a confronto

Adam

Combina momentum adattivo e scaling per coordinata. Converge velocemente, ma può oscillare.

Adagrad

Adatta il passo per coordinata accumulando i gradienti: aggiornamenti più stabili, ottimo per dati con scale diverse.

SVRG

Calcola periodicamente il gradiente completo per ridurre la varianza. Teoricamente solido, ma computazionalmente più costoso.

Risultati: Accuratezza vs. Epoche

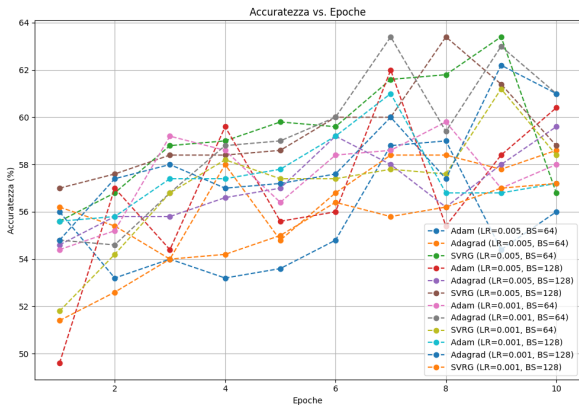


Figura: Confronto dell'accuratezza al variare delle epoche.

- Adam: convergenza rapida iniziale, oscillazioni nella loss.
- Adagrad: picchi di accuratezza elevati, buona stabilità.
- SVRG: comportamento stabile ma talvolta inferiore rispetto ad Adagrad.

Risultati: Loss vs. Epoche

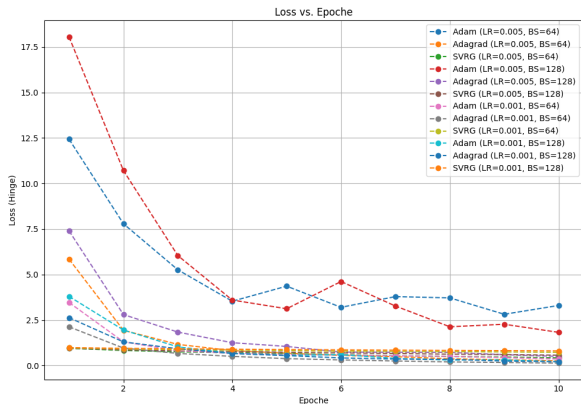


Figura: Andamento della Hinge Loss.

- Tutti gli ottimizzatori minimizzano la loss nel tempo.
- Adam mostra oscillazioni più marcate rispetto agli altri.

Risultati: Accuratezza vs. Tempo di addestramento

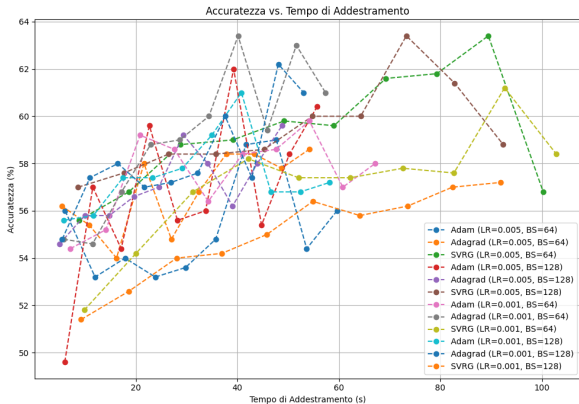


Figura: Accuratezza in funzione del tempo.

- SVRG è spesso più lento per epoca, ma può offrire stabilità.
- Adagrad tende ad essere efficiente e stabile su dati di feature scale diverse.
- Adam è spesso il più veloce per prototipi ma con possibili compromessi di stabilità.

- **Adam:**

- Convergenza rapida iniziale
- Accuratezza tipica: 55 %–60 %
- Oscillazioni nella loss

- **Adagrad:**

- Decrescita stabile della loss
- Accuratezza migliore: fino a 63 %–64 %
- Buon compromesso stabilità/prestazioni

- **SVRG:**

- Loss regolare, ma miglioramenti non sempre superiori
- Accuratezza simile o leggermente inferiore a Adagrad
- Tempo per epoca circa doppio

- ① La teoria non basta: SVRG ha buone garanzie ma non sempre rende meglio nella pratica.
- ② Gli algoritmi adattivi restano forti: Adagrad sorprendentemente competitivo.
- ③ Il contesto guida la scelta:
 - Adam: veloce, semplice, buono per prototipi.
 - Adagrad: stabile, ottimo per dati con feature di scale diverse.
 - SVRG: utile in problemi più grandi e con forte rumore nel gradiente.

Conclusioni

- Non esiste un ottimizzatore “migliore” in assoluto.
- La scelta dipende da:
 - Dimensione e natura del problema
 - Stabilità desiderata
 - Risorse computazionali
- La sperimentazione resta fondamentale per comprendere le interazioni tra modello, dati e ottimizzazione.