

IA - Progetto

Introduzione

Iper-parametri:

Hidden Channels

| Indica il numero di neuroni nei layer nascosti della rete neurale.

Valori alti :

- Migliorano la capacità della rete di catturare pattern complessi.
- Possono portare a una migliore generalizzazione se combinati con una buona regolarizzazione.

Valori bassi :

- Possono limitare la capacità di apprendimento, soprattutto con dati complessi.
- Tuttavia, possono prevenire l'overfitting su dataset più piccoli.

Learning Rate

| Controlla quanto velocemente il modello aggiorna i pesi durante l'addestramento.

Valori bassi :

- Consentono aggiornamenti più stabili e una convergenza più precisa.

Valori alti :

- Permettono aggiornamenti rapidi, ma possono causare oscillazioni e mancata convergenza.

Batch Size

| Indica quanti campioni vengono elaborati simultaneamente prima di aggiornare i pesi.

Batch size alti :

- Stabilizzano l'ottimizzazione e riducono la varianza tra gli aggiornamenti.
- Spesso migliorano la generalizzazione.

Batch size bassi :

- Consentono aggiornamenti più frequenti e adattamenti più rapidi, ma con maggiore rumore.

Num Neighbors

| Indica quanti vicini vengono considerati nel grafo per ogni nodo. Un valore più alto include più contesto.

Valori alti :

- Aiutano con grafi complessi, migliorando la qualità delle rappresentazioni apprese.

Valori bassi :

- Possono ridurre il rumore, ma rischiano di perdere informazioni importanti.

Negative Sampling Ratio

| Indica il numero di esempi negativi generati per ogni esempio positivo durante l'addestramento.

Valori alti :

- Aiutano quando il dataset è sbilanciato, fornendo più esempi negativi per un migliore apprendimento.

Valori bassi :

- Sono sufficienti se il dataset è già ben bilanciato.

Metriche di valutazione:

AUC

| Area sotto la curva ROC.

F1-score:

| Media armonica di precision e recall.

Precision:

| Numero di esempi positivi classificati correttamente diviso per il numero totale di esempi classificati come positivi

Recall:

Numero di esempi positivi classificati correttamente diviso per il numero totale di esempi positivi effettivi nel set di test

Loss:

Funzione di perdita.

Metodo per la scelta delle configurazioni

Poiché AUC e F1-score devono essere massimizzati mentre la Loss deve essere minimizzata, ho normalizzato i valori per renderli comparabili su una scala comune tra 0 e 1:

- $AUC_{\text{norm}} = \frac{AUC - AUC_{\min}}{AUC_{\max} - AUC_{\min}}$
- $F1_{\text{norm}} = \frac{F1 - F1_{\min}}{F1_{\max} - F1_{\min}}$
- $Loss_{\text{norm}} = \frac{Loss - Loss_{\min}}{Loss_{\max} - Loss_{\min}}$

La Loss è invertita perché vogliamo che valori più bassi siano considerati migliori.

1. Creazione di uno Score combinato

Per ogni configurazione, ho calcolato uno **Score finale** come combinazione ponderata delle metriche normalizzate:

$Score = 0.4 \cdot AUC_{\text{norm}} + 0.4 \cdot F1_{\text{norm}} + 0.2 \cdot Loss_{\text{norm}}$

Le pesature sono state scelte così:

- AUC (40%):** Indica la capacità del modello di distinguere tra classi positive e negative.
- F1-score (40%):** Misura il bilanciamento tra Precision e Recall, cruciale per modelli con dati sbilanciati.
- Loss (20%):** Considerata meno rilevante rispetto a AUC e F1-score, ma comunque indicativa della qualità dell'addestramento.

1. Selezione delle 5 migliori configurazioni

Dopo aver calcolato lo Score per tutte le configurazioni presenti nei file, ho selezionato le **5 configurazioni con il punteggio più alto**.

Motivazione della scelta dei criteri

1. Perché AUC e F1-score?

- L'AUC è una misura globale della qualità del modello. Un AUC più alto significa che il modello separa meglio le classi.
- L'F1-score bilancia Precision e Recall, quindi assicura che il modello non sia troppo sbilanciato verso falsi positivi o falsi negativi.

2. Perché includere la Loss?

- La Loss misura direttamente l'errore del modello. Anche se non sempre un minor valore di Loss significa migliori metriche di valutazione, aiuta a prevenire selezioni di modelli con ottime metriche ma addestramento instabile.

3. Perché pesi diversi?

- AUC e F1-score sono più importanti per la qualità del modello finale, quindi ricevono un peso maggiore (40% ciascuno).
- La Loss ha un peso minore (20%) perché, sebbene importante, non sempre è correlata direttamente alla performance finale in classificazione.

Analisi dei dati

Base

codice originale del prof cioè il link tra film e movie, completato il notebook colab e testato con le varie configurazioni

Optimizer: Adam

Hidden Channels	Learning Rate	Batch Size	Num Neighbors	Neg Sampling Ratio	AUC	F1-score	Precision	Recall	Loss
64	0.005	128	[20, 10]	1	0.931089	0.796777	0.809246	0.784687	0.0635
128	0.001	64	[30, 15]	1	0.931229	0.796395	0.799700	0.793117	0.0829
128	0.001	128	[30, 15]	2	0.935655	0.788131	0.829407	0.750769	0.0942
32	0.005	128	[30, 15]	1	0.931749	0.803575	0.791590	0.815928	0.1177
64	0.005	128	[30, 15]	2	0.930632	0.777145	0.847406	0.717644	0.0857

Analisi Adam:

Le configurazioni migliori tendono ad avere un numero maggiore di Hidden Channels (64-128) e un Learning Rate relativamente basso (0.001-0.005). La combinazione di Batch Size 128 e un maggiore numero di vicini ([30, 15]) sembra favorire risultati migliori. L'AUC rimane costantemente alto, sopra 0.93, mentre l'F1-score si mantiene attorno a 0.78-0.80.

Optimizer: AdamW

Hidden Channels	Learning Rate	Batch Size	Num Neighbors	Neg Sampling Ratio	AUC	F1-score	Precision	Recall	Loss
256	0.0005	128	[20, 10]	1	0.929335	0.798373	0.780017	0.817614	0.0765

128	0.0005	64	[20, 10]	2	0.940899	0.791337	0.851302	0.739264	0.1166
64	0.0050	64	[30, 15]	2	0.930821	0.787841	0.838652	0.742834	0.0842
128	0.0010	64	[30, 15]	1	0.929097	0.795486	0.790577	0.800456	0.0961
256	0.0010	128	[10, 5]	1	0.928361	0.775956	0.819526	0.736785	0.0746

Analisi AdamW:

Le migliori configurazioni utilizzano un numero elevato di Hidden Channels (128-256) e un Learning Rate molto basso (0.0005-0.001). Il numero di vicini varia leggermente, ma generalmente configurazioni con [20, 10] e [30, 15] sembrano avere performance elevate. L'AUC è generalmente superiore a 0.92, mentre l'F1-score si mantiene tra 0.77 e 0.80.

Optimizer: SGD

Hidden Channels	Learning Rate	Batch Size	Num Neighbors	Neg Sampling Ratio	AUC	F1-score	Precision	Recall	Loss
256	0.01	64	[20, 10]	1	0.911210	0.739666	0.787370	0.697411	0.2249
128	0.01	64	[20, 10]	1	0.899091	0.747698	0.738512	0.757116	0.2219
256	0.01	128	[30, 15]	1	0.892388	0.730655	0.728996	0.732322	0.2391
256	0.01	64	[10, 5]	3	0.905186	0.657782	0.839077	0.540910	0.2060
128	0.005	64	[10, 5]	1	0.879099	0.715174	0.711941	0.718437	0.2509

Analisi SGD:

A differenza di Adam e AdamW, le configurazioni ottimali con SGD tendono ad avere un Learning Rate più alto (0.01). L'uso di 256 Hidden Channels sembra produrre buoni risultati, mentre il numero di vicini varia. L'AUC è generalmente inferiore rispetto ad Adam e AdamW, con un massimo di 0.91. Anche l'F1-score è leggermente più basso, attorno a 0.73-0.75. Inoltre, il Loss è più elevato rispetto agli altri ottimizzatori.

Considerazioni Finali

L'analisi degli ottimizzatori mostra che **Adam e AdamW sono le scelte migliori** per la link prediction su MovieLens, con performance superiori a **SGD** in termini di **AUC e F1-score**.

Adam offre un miglior bilanciamento tra le metriche, mentre AdamW tende ad avere un AUC leggermente più alto. **SGD, invece, presenta una Loss più elevata e risultati peggiori**, suggerendo che potrebbe non essere l'ideale per questo compito.

Tra i parametri, **il numero di vicini (Num Neighbors) influenza le prestazioni**, con configurazioni come **[20,10] e [30,15]** che emergono come le più efficaci. Inoltre, **Batch Size 128** è ricorrente tra le migliori configurazioni, soprattutto con Adam e AdamW. **Adam e AdamW funzionano meglio con Hidden Channels elevati (128-256), un Learning Rate basso (0.0005-0.001) e Batch Size 128**, mentre **SGD preferisce un Learning Rate più alto (0.01) e può funzionare con meno Hidden Channels**.

Infine, **Num Neighbors e Negative Sampling Ratio influenzano le prestazioni**, con **[30,15] e un ratio di 1-2** che si rivelano parametri comuni tra le configurazioni migliori.

Estensione Base

Mettiamo soltanto nella base di conoscenza solo la tripla utente-rate-movie che viene accettata come feedback dall'utente con restrizioni che l'utente accetta soltanto il primo film consigliato e ne accetta un massimo di 5 film alla volta

Risultati per Adam

Hidden Channels	Learning Rate	Batch Size	Num Neighbors	Neg Sampling Ratio	AUC	F1-score	Precision	Recall	Loss
128	0.01	64	[10, 5]	1	0.8939	0.7393	0.7321	0.7467	0.1805
256	0.01	64	[30, 15]	1	0.9005	0.7438	0.6420	0.8838	0.2298
256	0.01	64	[10, 5]	1	0.9050	0.7452	0.7526	0.7381	0.2630
256	0.01	64	[20, 10]	3	0.9073	0.7267	0.8038	0.6631	0.2559
256	0.01	64	[20, 10]	2	0.9101	0.7689	0.7382	0.8022	0.3579

Considerazioni su Adam

Adam predilige **Hidden Channels più bassi (32-128) e un Learning Rate di 0.001 o inferiore**, indicando che un aggiornamento più controllato migliora la performance. Il **Batch Size di 64** rimane la scelta più frequente, e il **Neg Sampling Ratio tende a 1**, suggerendo che, a differenza di SGD, Adam non necessita di molti esempi negativi per generalizzare bene.

Risultati per AdamW

Hidden Channels	Learning Rate	Batch Size	Num Neighbors	Neg Sampling Ratio	AUC	F1-score	Precision	Recall	Loss
64	0.001	64	[30, 15]	2	0.9339	0.7660	0.8535	0.6947	0.0700
128	0.001	64	[30, 15]	2	0.9334	0.7823	0.8336	0.7369	0.1153
64	0.005	128	[30, 15]	1	0.9270	0.7946	0.7737	0.8166	0.1259
32	0.001	64	[20, 10]	1	0.9228	0.7877	0.7324	0.8521	0.1022
128	0.0005	64	[30, 15]	2	0.9302	0.7823	0.8142	0.7528	0.1218

Analisi AdamW

AdamW funziona meglio con **Learning Rate più bassi (0.0005 - 0.0010) e Batch Size di 64**, mostrando maggiore stabilità rispetto a SGD. Le configurazioni migliori ottengono **AUC e F1-score più alti**, suggerendo che AdamW possa essere più efficace nel bilanciare precisione e richiamo. Il **Neg Sampling Ratio intorno a 2** sembra fornire buoni risultati, evitando sia il sovracampionamento che la scarsità di esempi negativi.

Risultati per SGD

Hidden Channels	Learning Rate	Batch Size	Num Neighbors	Neg Sampling Ratio	AUC	F1-score	Precision	Recall	Loss
128	0.01	64	[10, 5]	1	0.8939	0.7393	0.7321	0.7467	0.1805
256	0.01	64	[30, 15]	1	0.9005	0.7438	0.6420	0.8838	0.2298
256	0.01	64	[10, 5]	1	0.9050	0.7452	0.7526	0.7381	0.2630
256	0.01	64	[20, 10]	3	0.9073	0.7267	0.8038	0.6631	0.2559
256	0.01	64	[20, 10]	2	0.9101	0.7689	0.7382	0.8022	0.3579

Analisi SGD

Le migliori configurazioni per **SGD** mostrano che **un alto numero di Hidden Channels (128-256) e un Learning Rate di 0.01** sono scelte ottimali. Il **Batch Size di 64** appare come un buon compromesso. Il **Neg Sampling Ratio tra 1 e 3** suggerisce che un moderato numero di esempi negativi aiuta il modello senza peggiorarne la stabilità. Non sempre la Loss più bassa corrisponde alle migliori prestazioni, confermando che AUC e F1-score sono metriche più rilevanti.

Considerazioni finali

AdamW e Adam tendono a essere più stabili di SGD, con AdamW che ottiene il punteggio più alto complessivo. **SGD funziona meglio con Learning Rate più alti e Hidden Channels elevati**, mentre **Adam e AdamW preferiscono Learning Rate più bassi e Hidden Channels più contenuti. Il Batch Size ottimale è quasi sempre 64. Il Neg Sampling Ratio ideale varia:** SGD beneficia di valori più alti, mentre Adam e AdamW lavorano meglio con valori più bassi.

Estensione Avanzata

| L'utente accetta i primi 2 film consigliati con un massimo di 20

Adam

Hidden Channels	Learning Rate	Batch Size	Num Neighbors	Neg Sampling Ratio	AUC	F1-score	Precision	Recall	Loss
128	0.0005	64	[20, 10]	1	0.930299	0.787891	0.803277	0.773083	0.053287
128	0.0010	128	[20, 10]	1	0.931001	0.797550	0.794566	0.800555	0.098145
32	0.0050	128	[20, 10]	1	0.932942	0.804045	0.781502	0.827928	0.118295
128	0.0010	64	[20, 10]	1	0.930221	0.797665	0.782667	0.813250	0.117257
32	0.0050	64	[20, 10]	2	0.936738	0.801292	0.829354	0.775067	0.160618

Analisi Adam

Adam si conferma l'ottimizzatore più **bilanciato e stabile**, con valori di **AUC e F1-score costantemente alti** e una Loss generalmente più bassa rispetto agli altri. Funziona bene con un **learning rate tra 0.0005 e 0.005**, hidden channels sia bassi (32) che alti (128), e batch size di **64 o 128. Il Negative Sampling Ratio = 1** è sufficiente nella maggior parte dei casi, anche se in alcuni scenari **valori più alti migliorano l'AUC**.

AdamW

Hidden Channels	Learning Rate	Batch Size	Num Neighbors	Neg Sampling Ratio	AUC	F1-score	Precision	Recall	Loss
32	0.0010	64	[20, 10]	1	0.931510	0.796493	0.782617	0.810870	0.065112
32	0.0050	128	[20, 10]	1	0.931170	0.802288	0.772323	0.834672	0.102286
128	0.0010	64	[30, 15]	2	0.935746	0.790437	0.836378	0.749281	0.116591
128	0.0010	128	[30, 15]	3	0.932025	0.765351	0.851056	0.695329	0.097609
128	0.0010	64	[20, 10]	1	0.932876	0.802198	0.808091	0.796390	0.177580

Analisi AdamW

AdamW mostra un comportamento simile ad Adam ma con un **learning rate più basso (0.001 - 0.005)** e una maggiore variabilità nei parametri ottimali. Risulta più efficace con **hidden channels tra 32 e 128**, con **batch size 64 o 128** e **Num Neighbors più flessibile ([20,10] o [30,15])**. Alcune configurazioni ottimali sfruttano **Negative Sampling Ratio = 2 o 3**, suggerendo che una maggiore quantità di esempi negativi possa migliorare le prestazioni in alcuni contesti.

SGD

Hidden Channels	Learning Rate	Batch Size	Num Neighbors	Neg Sampling Ratio	AUC	F1-score	Precision	Recall	Loss
256	0.0050	64	[30, 15]	1	0.890785	0.738877	0.674445	0.816920	0.178487
256	0.0100	128	[20, 10]	1	0.891863	0.741893	0.703207	0.785084	0.208460
256	0.0100	64	[20, 10]	2	0.909809	0.713752	0.814919	0.634930	0.235853
256	0.0100	64	[30, 15]	3	0.904337	0.700564	0.812754	0.615591	0.244844
128	0.0100	64	[30, 15]	1	0.895615	0.733649	0.736990	0.730338	0.292228

Analisi SGD

SGD ha un **AUC inferiore rispetto ad Adam e AdamW**, ma le sue migliori configurazioni puntano su **hidden channels elevati (128-256) e learning rate più alti (0.005 - 0.01)**. Batch size **64 o 128** funzionano meglio, mentre il Negative Sampling Ratio varia tra **1 e 3**. Questo suggerisce che, per competere con Adam, SGD ha bisogno di un modello più complesso e di un apprendimento più aggressivo.

Considerazioni Finali

Adam e AdamW sono gli ottimizzatori migliori per la link prediction su MovieLens, garantendo **stabilità e alte performance**. Adam è il più bilanciato, con buoni risultati già con **Negative Sampling Ratio = 1**, mentre AdamW è più sensibile agli esempi negativi e beneficia di **valori tra 2 e 3**. SGD, invece, richiede **hidden channels elevati e learning rate più alti** per ottenere risultati competitivi, ma resta meno efficace rispetto ad Adam e AdamW.

Estensione Dinamica

Viene aggiornato oltre che la tripla user-rate-movie anche la feature utente dello user con i nuovi film consigliati

Adam

Hidden Channels	Learning Rate	Batch Size	Num Neighbors	Neg Sampling Ratio	AUC	F1-score	Precision	Recall	Loss
64	0.005	128	[20, 10]	1	0.9307	0.8043	0.7685	0.8437	0.1115
128	0.001	64	[20, 10]	1	0.9322	0.8086	0.7758	0.8443	0.2388
64	0.005	128	[20, 10]	2	0.9343	0.8001	0.8125	0.7882	0.2072
256	0.001	128	[20, 10]	1	0.9291	0.8015	0.7857	0.8179	0.1455
128	0.001	128	[20, 10]	1	0.9308	0.7987	0.7846	0.8133	0.1893

Considerazioni:

- Le migliori configurazioni utilizzano **batch size medio-grande (64-128)** e **learning rate più basso (0.001-0.005)**, indicando stabilità nell’addestramento.
- Numero di canali tra 64 e 256**, suggerendo che reti più grandi possono migliorare le prestazioni.
- Sampling Ratio = 1 o 2**, mostrando che un moderato bilanciamento nei dati negativi è vantaggioso.

AdamW

Hidden Channels	Learning Rate	Batch Size	Num Neighbors	Neg Sampling Ratio	AUC	F1-score	Precision	Recall	Loss
128	0.0005	64	[20, 10]	2	0.9347	0.7839	0.8369	0.7373	0.0893
64	0.0050	128	[30, 15]	1	0.9272	0.7997	0.7648	0.8379	0.0936
64	0.0010	64	[20, 10]	1	0.9303	0.8013	0.7714	0.8337	0.1100
32	0.0100	128	[10, 5]	1	0.9235	0.7888	0.7704	0.8082	0.0882
32	0.0050	64	[20, 10]	2	0.9331	0.7999	0.8073	0.7925	0.1612

Considerazioni:

- AdamW favorisce **learning rate molto basso (0.0005-0.005)**, il che indica una convergenza più stabile.
- Le reti con **64-128 hidden channels** si comportano meglio.
- Anche qui, **sampling ratio tra 1 e 2** è ottimale.
- Il **batch size varia tra 64 e 128**, senza una chiara preferenza.

SGD

Hidden Channels	Learning Rate	Batch Size	Num Neighbors	Neg Sampling Ratio	AUC	F1-score	Precision	Recall	Loss
256	0.005	64	[30, 15]	1	0.8930	0.7382	0.7179	0.7598	0.1733
128	0.010	64	[30, 15]	1	0.8981	0.7509	0.6936	0.8185	0.2299
256	0.010	64	[20, 10]	1	0.9097	0.7673	0.6940	0.8580	0.3098
256	0.010	128	[10, 5]	1	0.8882	0.7389	0.6866	0.7998	0.1975
128	0.010	64	[20, 10]	1	0.8999	0.7349	0.7550	0.7158	0.2379

Considerazioni:

- Rispetto agli altri ottimizzatori, **SGD ottiene punteggi inferiori** su AUC e F1-score, suggerendo che è meno efficace in questo contesto.
- Learning rate più alto (0.005-0.010)** rispetto ad Adam e AdamW, il che è coerente con la natura più "semplice" di SGD.
- Hidden Channels tra 128 e 256** performano meglio, confermando che reti più profonde sono vantaggiose.
- Batch Size più piccolo (64) tende a funzionare meglio**, tranne in un caso (128).

Conclusioni Generali

- Adam** e **AdamW** performano meglio rispetto a **SGD**, in particolare in termini di AUC e F1-score.
- AdamW** richiede learning rate più basso rispetto ad Adam, ma le configurazioni migliori sono simili.
- SGD fatica a raggiungere gli stessi livelli di performance**, indicando che potrebbe non essere l'ottimizzatore ideale per questo task.
- Hidden Channels tra 64 e 256** e **Batch Size tra 64 e 128** sembrano essere un buon compromesso in tutti i casi.
- Neg Sampling Ratio tra 1 e 2** è preferibile, evitando sia un sottocampionamento che un eccessivo bilanciamento.

Conclusione

L'analisi ha dimostrato che Adam e AdamW sono i migliori ottimizzatori, fornendo prestazioni superiori rispetto a SGD, che ha mostrato un’ AUC più bassa e una loss maggiore.

Adam è più stabile e bilanciato poiché funziona al meglio con Hidden Channel compresi tra 32 e 128, Learning Rate compreso tra 0,0005 e 0,005 e Batch Size di 64 o 128.

Inoltre, un Neg Sampling Ratio di 1 è generalmente sufficiente per ottenere buoni risultati.

AdamW, invece, mostra una maggiore sensibilità agli esempi negativi e ottiene i migliori risultati con un Neg Sampling Ratio compreso tra 2 e 3.

SGD non si è comportato altrettanto bene: sebbene possa lavorare con Hidden Channels più alti (128-256) e un Learning Rate maggiore (0,005-0,01), i suoi risultati rimangono inferiori a quelli ottenuti con Adam e AdamW.

Anche altri parametri hanno influenzato le prestazioni.

Il Number of Neighbors si è rivelato fondamentale, con le configurazioni (20,10) e (30,15) che hanno fornito risultati più soddisfacenti.

La dimensione ottimale del batch è 64 o 128, con una leggera preferenza per il valore più alto.

In conclusione, Adam è l'opzione più equilibrata, mentre AdamW è più adatto a contesti in cui la gestione di esempi negativi è fondamentale. SGD rimane meno efficace, richiedendo più risorse senza garantire prestazioni comparabili. La combinazione vincente comprende un basso Learning rate, Hidden Channels tra 32 e 128 e un adeguato Neg Sampling Ratio.