Large Language Models

Architettura, Addestramento e Impatti Applicativi

[Il tuo nome]

Anno Accademico 2024–2025

Indice

Capitolo 1

Introduzione

1.1 Introduzione

Contesto scientifico

Motivazioni

Obiettivi della tesi

Breve descrizione dei capitoli

Capitolo 2

Autoencoders

2.1 Introduzione

Per comprendere le metodologie adottate in questo lavoro di tesi è necessario introdurre una particolare classe di architetture di reti neurali, note come *autoencoder*. Gli autoencoder sono modelli di apprendimento non supervisionato progettati per apprendere una rappresentazione compatta dei dati di input, attraverso un processo di compressione e successiva ricostruzione. Le prime formulazioni di questo approccio risalgono ai lavori di Rumelhart, Hinton e Williams alla fine degli anni '80 [**hinton1987autoencoders**].

L'idea di base di un autoencoder consiste nell'addestrare una rete neurale a mappare i dati di input in uno spazio latente di dimensione ridotta, per poi ricostruire i dati originali a partire da tale rappresentazione. Le variabili che descrivono lo spazio intermedio sono comunemente indicate come *variabili latenti* e costituiscono una descrizione compressa ma informativa dei dati di partenza.

Si consideri un dataset di addestramento S_T costituito da M osservazioni non etichettate x_i, con $i = 1, \ldots, M$:

$$S_T = \{x_1, x_2, \ldots, x_M\}. \tag{2.1}$$

In generale, ciascuna osservazione appartiene allo spazio $\mathbb{R}^N$, ovvero $x_i \in \mathbb{R}^N$. L'obiettivo dell'autoencoder è quello di apprendere una funzione che permetta di ricostruire ciascun dato di input nel modo più accurato possibile, minimizzando una misura dell'errore di ricostruzione tra l'input originale e l'output della rete. Come mai si può essere interessati a questo tipo di

operazione? Per rispondere a questa domanda si fornisce di seguito una definizione

Definition 1. *Un autoencoder è un tipo di algoritmo il cui scopo principale è apprendere una rappresentazione dei dati, utilizzabile per diverse applicazioni, imparando a ricostruire in modo sufficientemente accurato un insieme di osservazioni di input [1].*

Per comprendere meglio il funzionamento degli autoencoder, si consideri la Figura 2.1.

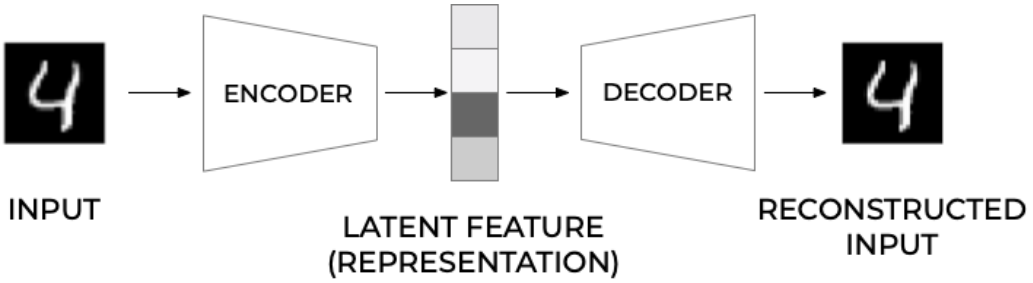

Figura 2.1: Schema di funzionamento di un autoencoder [2]

Bibliografia

[1] Dor Bank, Noam Koenigstein e Raja Giryes. "Autoencoders". In: *CoRR* abs/2003.05991 (2020). arXiv: 2003.05991. URL: https://arxiv.org/abs/2003.05991.

[2] Umberto Michelucci. *An Introduction to Autoencoders*. 2022. arXiv: 2201.03898 [cs.LG]. URL: https://arxiv.org/abs/2201.03898.