SOMMARIO

Tecniche di Machine Learning per analisi astrofisiche

Le più recenti generazioni strumenti di osservazione astronomica producono una enorme quantità di dati che necessita di una robusta ed efficiente analisi. Per questo motivo, nel campo dell'analisi di dati si utilizzano metodi e tecnologie di analisi di big data. Tra queste, gli algoritmi di Machine Learning (ML) hanno ottenuto grande successo nel campo astronomico e astrofisico. Questi strumenti sono diventati strumenti fondamentali per gestire e interpretare l'enorme quantità di dati prodotti, fornendo nuovi insights nei campi dell'astrofisica.

Suddivisione in Supervised e Unsupervised

Gli algoritmi di ML si suddividono in due categorie principali: Supervised ML e Unsupervised ML. I primi vengono utilizzati per modellizzare una mappatura tra un insieme di proprietà e un set di variabili target. Questo modello si basa sull'apprendimento compiuto su coppie di variabili di input e di output fornite da chi opera l'analisi tramite esempi di simulazioni o di dati reali. I secondi, gli algoritmi Unsupervised, vengono maggiormente utilizzati per scopi di clustering, diminuzione di dimensionalità e rilevamento di anomalie in dataset non etichettati.

Applicazioni

- Classificazione di immagini:

l'obiettivo è quello di classificare il contenuto delle immagini sulla base di etichette predefinite e in quanto tale rientra negli algoritmi di ML.

- Identificazione di sorgente:

è il processo di identificazione di una sorgente astrofisica da una immagine, in cui l'algoritmo valuta la presenza di una deviazione significativa dei valori dei pixel dal livello del background.

- Localizzazione di sorgente:

L'identificazione di una sorgente in un'immagine permette di valutare alcune sue proprietà, tra cui la posizione.

- Deblending:

Il deblending, o scomposizione, si riferisce all'identificazione dei diversi contributi ad un unico segnale, ad esempio, nell'immagine in cui la luce di due sorgenti sono sovrapposte.

Supervised

La differenza principale tra i modelli di Machine Learning Supervisionato (ML) e le tradizionali tecniche di adattamento del modello è che, in quest'ultimo caso, il modello è

predefinito, mentre nel Machine Learning il modello viene costruito sulla base dei nuovi dati.
Esistono tre fasi nell'applicazione di un algoritmo Supervised ML. Nella fase di addestramento, un sottoinsieme del dataset di input, chiamato set di training, viene utilizzato per "addestrare" il modello, costruendo i parametri necessari per descriverlo. Nella successiva fase di convalida, i parametri del modello vengono ottimizzati attraverso successive iterazioni utilizzando una funzione di costo (Loss function), che misura la bontà delle previsioni rispetto alla classificazione reale. Questo processo continua fino a trovare il set di parametri che meglio si adatta a descrivere il set di convalida. Nella fase finale di test, il modello viene utilizzato per fare previsioni su un set di test al fine di valutare le prestazioni complessive del modello, magari confrontando diverse implementazioni di algoritmi di Machine Learning.

Esistono tanti algoritmi di Supervised ML, che vengono impiegati in base al caso d'uso. Gli algoritmi principali sono il Support Vector Machine, Random Forests, e Artificial Neural Networks.

Support Vector Machine

Il Support Vector Machine (SVM) è un algoritmo di apprendimento supervisionato utilizzato per la regressione di dati e la classificazione di oggetti con un alto numero di caratteristiche. Dato un dataset con vettori composti di N caratteristiche, l'algoritmo di Support Vector Machine si occupa di trovare il miglior iperpiano che suddivide l'iperpiano N-dimensionale e le N caratteristiche. Una volta che l'algoritmo ha stabilito il miglior iperpiano, questo viene utilizzato su nuovi valori per decidere l'appartenenza di questi a una classe o a un'altra.

Decision Trees

Si tratta di un algoritmo supervisionato utilizzato principalmente per la classificazione. È un modello decisionale ad albero, in cui ogni ramificazione è basata sul controllo di una condizione su una caratteristica del dataset. Viene utilizzato per classificazione e regressione di dati. I Decision Trees sono un modo per analizzare le proprietà di input per prevedere un possibile outcome, ad esempio le caratteristiche fisiologiche di un individuo con la sua probabilità di sopravvivenza, o per distinguere stelle da galassie.

Random Forests

L'algoritmo di Random Forest è una collezione di Decision Trees, dove diversi Decision Trees vengono allenati su subset diversi del dataset complessivo, considerando subset diversi delle features che caratterizzano il dataset. La previsione del Random Forest risulta essere la classe che è stata prevista da più decision trees, ossia, l'oggetto viene esaminato dai vari decision trees e la classe con più "voti" è il risultato della Random Forest. Le Random Forest non tengono conto però delle incertezze relative al risultato.

Nelle Probabilistic Random Forest (PRF) si introduce una funzione di distribuzione di probabilità, in cui il valore atteso è la misura data e la varianza è l'incertezza.
L'algoritmo di Probabilistic Random Forest è stato utilizzato in [Guarneri 2021 - The probabilistic random forest applied to the selection of quasar candidates in the QUBRICS survey] per identificare nuovi QSO nell'emisfero meridionale. I dati fotometrici ottenuti dai catalogh SkyMapper DR1, Gaia DR2, 2MASS, WISE e GALEX vengono valutati con questo algoritmo di classificazione e più di 600 nuovi QSO candidati sono risultati dal modello di classificazione.

Artificial Neural Networks

Le Artificial Neural Networks sono algoritmi con strutture versatili, ispirate dalla biologia delle reti neurali del cervello. Sono strutture che consistono in strati in cui le informazioni si propagano dallo strato iniziale che contiene gli input a quello finale che contiene il risultato. Tra questi due strati si trovano strati intermedi, o nascosti (hidden layers). Ogni strato è composto da nodi, o neuroni, attraverso cui l'informazione contenuta nei nodi di input viene scomposta e analizzata in maniera sequenziale, portando alla ricostruzione e al riconoscimento di features di basso livello negli strati nascosti iniziali, fino agli strati finali, in cui queste features costruiscono caratteristiche di più alto livello. Questo può essere descritto bene nel caso in cui l'input è l'immagine di un oggetto e il task della rete neurale è la classificazione dell'oggetto. Ogni nodo di input è rappresentato dai pixel dell'immagine. Lo strato successivo "cattura" informazioni di basso livello, come ad esempio bordi delle immagini. Gli strati successivi operano su questi bordi, fino a riconoscere, ad esempio, la forma dell'oggetto. Queste informazioni si propagano fino allo strato di output, in cui l'oggetto viene classificato. Grazie a questa struttura, le reti neurali artificiali si prestano a numerosi task, quali clustering, classificazione, regressione e altri.

Feed Forward Neural Network

Tra questi, il Feed Forward Neural Network (FFNN) è un algoritmo di ML che può essere impiegato come estimator: è un processo di apprendimento in cui l'informazione viaggia in maniera unidirezionale, dallo strato di input attraverso i layer nascosti, fino all'output. Coinvolge l'adattamento del modello ai dati di addestramento in modo che possa generalizzare a nuovi dati non osservati, facendo previsioni accurate. Si tratta di una generalizzazione della regressione logistica, in cui l'algoritmo può apprendere pattern più complessi incorporando più strati nascosti di neuroni con funzioni di attivazione non lineare.
Nel suo utilizzo in [Crupi 2023 - Searching for long faint astronomical high energy transients: a data driven approach], questo algoritmo viene usato per stimare il valore di background osservativo dei satelliti HERMES pathfinder sulla base delle variabili e proprietà dei satelliti (posizione in orbita, inclinazione, esposizione al sole ecc). Il software di Background Estimator viene allenato e testato nei dati del Fermi GBM. Il valore così ottenuto con il modello del background viene confrontato con il valore osservativo e l'eccesso viene quantificato in unità di deviazioni standard, tramite la

tecnica di Poisson-FOCuS, in modo da poter definire una soglia oltre la quale si può identificare un eccesso nel count rate osservato.

Recurrent Neural Network

Al contrario delle reti neurali Feed Forward, le reti neurali ricorrenti (RNN) hanno la possibilità di mantenere la memoria delle informazioni precedenti introducendo una connessione ricorrente, in cui quindi un loop si viene a creare da un neurone verso sé stesso, o verso neuroni dello stesso strato. Questo permette alle RNN di “ricordare”, e quindi di catturare caratteristiche a lungo termine, o pattern dinamici. Sono quindi ideali per l’analisi di serie temporali.

Sono state ad esempio utilizzate nel lavoro di [Finke 2020 - Classification of Fermi-LAT sources with deep learning using energy and time spectra], in cui per studiare le sorgenti più incerte del catalogo si fa uso di una forma ibrida di reti neurali, usando una rete neurale densa, o fully connected (un tipo di FFNN), e una RNN. L’idea è di utilizzare questo modello per classificare sorgenti di raggi gamma non identificate precedentemente.

Le RNN sono state utilizzate anche in [Sadeh 2019 - Deep learning detection of transients] con lo scopo di produrre un modello capace di identificare lampi di raggi gamma fortuiti nel campo visivo dell’osservatorio del Cherenkov Telescope Array.

Convolutional Neural Network

Una Convolutional Neural Network (CNN) è un tipo di rete neurale Feed Forward. Una semplice FFNN è agnostica alle regolarità degli input, ma se si conoscono le caratteristiche a priori dei dati, possono essere introdotti dei vincoli nella rete neurale, in modo da poter ridurre il numero di parametri. Questo può essere fatto aggiungendo degli strati che manipolano i dati, tramite tecniche di convoluzione e di pooling, ossia filtri (kernels) applicati ai dati di input. Questo tipo di rete neurale viene principalmente utilizzata per l’analisi di immagini. I filtri convoluzionali vengono usati per mettere in risalto pattern e caratteristiche locali (bordi, simmetrie, forme), che quindi restituiscono delle mappe di features. Lo strato di pooling (vedi RoI Pooling) applica un filtro per ridurre le dimensioni, in particolare può essere applicato un pooling di massimo, che estrae il valore massimo da una certa regione (esempio, una sottogriglia 3x3 dell’immagine), oppure un pooling di media, che restituisce il valore medio di quella regione. Questa diminuzione dei parametri rende questa rete neurale particolarmente efficiente e performante a livello computazionale.
Questa rete è stata utilizzata nel lavoro di [Parmiggiani 2023 - A Deep-learning Anomaly-detection Method to Identify Gamma-Ray Bursts in the Ratemeters of the AGILE Anticoincidence System] per analisi di transienti astrofisici nelle serie temporali dei segnali gamma del satellite AGILE (Astro-rivelatore Gamma a Immagini Leggero), alla ricerca di lampi di raggi gamma. In particolare, i dati che vengono usati arrivano dal sistema di anticoincidenza del satellite. Questa rete neurale viene utilizzata per costruire

un Convolutional Autoencoder (vedi gli Autoencoder nel paragrafo Unsupervised), ossia una rete costituita da uno strato che decostruisce l'input, riduce le dimensioni mantenendo le caratteristiche più rilevanti, e uno strato che ricostruisce l'input.
Anche nel lavoro di [Parmiggiani 2021 - A Deep Learning Method for AGILE-GRID Gamma-Ray Burst Detection] è stata utilizzata una CNN per analizzare immagini delle mappe di intensità dello stesso satellite AGILE alla ricerca di lampi di raggi gamma.

Transformer

Un trasformatore è un tipo di architettura specializzabile in task che manipolano oggetti suddivisibili in sequenze, come testi nel caso del processing di linguaggio naturale e la traduzione automatica di testi. Il trasformatore ha un meccanismo chiamato di auto-attenzione, utilizzato per pesare diversamente le parti della sequenza di input durante la previsione, concentrandosi quindi su elementi rilevanti in modo dinamico. Altra caratteristica importante del trasformatore è la possibilità di elaborare una intera sequenza in parallelo, ossia il meccanismo di auto-attenzione viene eseguito su più "teste" (si parla infatti di multi-head attention), ogni testa con i propri parametri di trasformatore, il ché consente di apprendere e sperimentare diverse rappresentazioni di dipendenze tra gli elementi della sequenza. L'obiettivo del trasformatore è quello di prevedere il successivo valore di una sequenza, dove l'output di previsione sarà una la distribuzione di probabilità rispetto a tutti i possibili valori di sequenza, che, aggiunti in maniera iterativa, portano alla generazione di una nuova serie di sequenze.

Residual Neural Network

Una rete neurale residuale (ResNet) è un tipo di architettura che fa uso di collegamenti residui (detti shortcut, o skipconnections) che consentono alla rete di propagare i gradienti più efficacemente e di apprendere mapping residui, ossia le differenze tra l'output desiderato e l'output corrente di uno stato. In questo modo, la ResNet ha la possibilità di concentrarsi sui dettagli, invece di apprendere la mappa in maniera troppo generica. L'architettura è composta da una serie di blocchi residui che consistono in strati convoluzionali e skipconnection.
Le ResNet sono state utilizzate in compiti di visione artificiale, quindi classificazione di immagini e rilevazione e segmentazione di oggetti nelle immagini.

R-CNN

Le Region-based CNN (R-CNN) sono una famiglia di algoritmi di rilevamento degli oggetti in cui vengono analizzate regioni di immagini tramite le CNN. La proposta delle regioni avviene tramite una ricerca selettiva, un algoritmo che genera molteplici regioni candidate in un'immagine. Una volta analizzate dalle CNN, dalle regioni vengono estratti dei vettori di caratteristiche, un insieme di numeri che rappresenta le informazioni visive contenute nella regione, come colori, forme, pattern ecc. Questo vettore viene poi passato a delle Support Vector Machine per classificare l'oggetto e prevedere un riquadro che lo delimita.

Fast R-CNN

Si tratta di un miglioramento del modello R-CNN, in cui l'intera immagine viene passata al CNN per generare la mappa di vettori di caratteristiche e il passaggio di ricerca selettiva delle regioni è invece integrato nella rete stessa. Ogni proposta di regione è associata a una finestra di dimensioni fisse e le caratteristiche vengono estratte da questa finestra utilizzando la tecnica di pooling Region of Interest (RoI), che permette di condividere le caratteristiche tra tutte le proposte di regione, rendendo il processo più veloce e efficiente. Allo stesso modo delle R-CNN, vengono usate le SVM per classificare ogni proposta di regione e valutare se contiene o meno un oggetto.

RoI Pooling

Il Region of Interest (RoI) Pooling è un metodo utilizzato in particolare nelle CNN e nelle R-CNN per l'identificazione di oggetti. Il RoI pooling consiste nel ridurre la dimensionalità di un'immagine o di una mappa di caratteristiche senza perderne le informazioni. Questa manipolazione della dimensionalità viene effettuata per permettere una integrazione precisa tra uno strato e l'altro, mantenendo quindi gli input delle dimensioni consistenti per la rete in questione. Infatti, se uno strato è pronto ad accogliere un input di dimensione NxM, ma l'input reale ha un'altra dimensione, la rete neurale non può analizzare quell'input. Per l'estrazione delle informazioni salienti, il RoI pooling sceglie delle sottoregioni, o celle, di dimensioni adeguate, ne effettua il pooling e restituisce la mappa risultante da questo pooling, che avrà le dimensioni corrette affinché possa essere analizzato dallo strato della rete neurale.
Il tipo di pooling effettuato può essere un filtro che restituisce la somma dei valori della regione, il massimo dei valori, oppure anche la media.

Unsupervised

Al contrario delle tecniche di Machine Learning supervisionate, le tecniche non supervisionate non si affidano a dati etichettati, ossia viene a mancare una corrispondenza tra l'input e il valore atteso dei parametri che descrivono l'input e quindi modelli di questo tipo non possono essere addestrati con una descrizione della vera natura del problema. L'obiettivo delle tecniche non supervisionate è quello di rivelare pattern nascosti o strutture nei dati analizzati.
Le tecniche supervisionate e non supervisionate però non sono completamente distinte, in quanto possono essere utilizzate in maniera complementare in un unico algoritmo, ad esempio prendendo le caratteristiche ottenute dalle tecniche non supervisionate e analizzarle con tecniche supervisionate (è il caso del Convolutional Autoencoder).

Algoritmi di Clustering

Gli algoritmi di clustering si basano sul raggruppamento degli oggetti nel campione di input sulla base delle proprietà che caratterizzano gli oggetti stessi. In tal modo, gli oggetti vengono catalogati come appartenenti a cluster, ossia gruppi di oggetti che si identificano per la somiglianza di alcune proprietà. La definizione di un cluster è dettata dal criterio applicato in un determinato algoritmo di clustering.

La maggior parte degli algoritmi parte dal calcolo della distanza euclidea tra gli oggetti nello spazio euclideo delle loro singole proprietà. Il problema del calcolo della distanza è che le unità di misura scelte nel vettore delle caratteristiche dell'oggetto potrebbero non essere omogenee. È il caso delle osservazioni di galassie e lo studio di spettri, di curve di luce e di proprietà intrinseche alle galassie (quali massa, temperatura, ecc), in cui la distanza è sensibile alle differenti scale delle unità di misura. Una soluzione sarebbe quella di normalizzare i valori delle caratteristiche, oppure standardizzare e riscalare i valori di ogni parametro in modo che abbia media zero e deviazione standard uno.

I principali algoritmi sono:
- k-means: dove k è il numero di cluster in cui si vuole suddividere il campione. Il clustering viene effettuato partendo dal calcolo della distanza euclidea tra i vari oggetti. Ottenute le distanze, l'obiettivo è di ottenere i cluster il cui centroide minimizza la distanza degli oggetti da questo, ossia in modo da minimizzare la varianza all'interno del cluster per tutti parametri del vettore delle caratteristiche. Gli oggetti appartenenti ai cluster vengono inizialmente selezionati in maniera randomica e gli oggetti cambiano fino a che le distanze non siano minimizzate. Il parametro fondamentale è il numero di cluster k e la determinazione del numero di cluster può non essere banale e dipende dalla distribuzione degli oggetti secondo le varie feature.
- hierarchical clustering: è una tecnica di clustering in cui i dati vengono organizzati in cluster in maniera gerarchica. Si parte anche qui dal calcolo delle distanze euclidee e ogni dato rappresenta un cluster. Successivamente, e in maniera ricorsiva, i cluster più vicini vengono uniti, andando a creare nuovi cluster. Così, la formazione dei cluster continua fino a quando la distanza tra i cluster non supera una soglia predefinita. Il clustering può anche avvenire con il processo inverso, ossia in maniera divisiva, con i dati facenti parte di un unico cluster iniziale, che si suddivide iterativamente.

Algoritmi di riduzione della dimensionalità

Gli algoritmi di riduzione della dimensionalità hanno l'obiettivo di estrarre dai dati le caratteristiche più importanti, in modo da diminuire le proprietà (dimensioni) da analizzare, mantenendo comunque le informazioni più rilevanti.

Le tecniche più comuni sono:
- Principal Component Analysis (PCA): è una trasformazione o proiezione delle variabili dei dati in un nuovo sistema di variabili con dimensionalità ridotta. Le nuove variabili sono definite dalle componenti principali con la massima varianza nei dati e sono quindi combinazioni lineari delle variabili iniziali. La tecnica PCA non è altro che il calcolo della matrice di covarianza dei dati iniziali, dove gli autovettori sono le nuove componenti principali.
- t-Distributed Stochastic Neighbor Embedding (tSNE): è una tecnica di riduzione non lineare che incorpora i dati iniziali in uno spazio di due o tre dimensioni. In questo modo, gli oggetti con caratteristiche simili si troveranno vicini in questo spazio, mentre oggetti diversi saranno distanti. Per iniziare, per ciascuna coppia di oggetti, vengono calcolate le probabilità condizionate di similarità tra i punti. Queste probabilità misurano la somiglianza tra i punti nello spazio originale. La distribuzione di probabilità viene calcolata con un kernel gaussiano dipendente dalle distanze tra i punti e da un valore di scala chiamato perplexity. Allo stesso

modo, viene calcolata la distribuzione di probabilità nello spazio ridotto e successivamente l'ottimizzazione mira a minimizzare la divergenza tra le distribuzioni di probabilità nei due spazi, ottimizzando la posizione dei punti nello spazio ridotto. È un algoritmo molto utile soprattutto nello studio di dataset i cui oggetti appartengono a gruppi differenti, ossia cluster. Il parametro perplexity influisce sulla distribuzione delle probabilità di similarità e può influenzare i risultati. Rappresenta quindi la scala dei cluster a cui il tSNE è sensibile. Per una perplessità alta, i dati saranno generalmente appartenenti a pochi cluster, ossia porta a una analisi più globale. Al contrario, una perplessità bassa porta alla suddivisione in più cluster, anche quando i dati non appartengono a sottogruppi separati.

Autoencoder

Un autoencoder è un tipo di rete neurale non supervisionata che compie due operazioni sui dati di input, suddivise in due fasi, una di encoding e una di decoding. Durante la fase di encoding, il codificatore riduce i dati a una rappresentazione latente, chiamata codice o embedding, di dimensione inferiore, in modo da mantenere le informazioni più rilevanti, ma in forma compatta. Nella successiva fase di decoding, il decodificatore ricostruisce il dato iniziale a partire dalla versione compatta, in modo da ridurre le differenze tra il dato originale e quello finale.
La caratteristica principale dell'autoencoder è quindi quella di poter lavorare in maniera performante sulle caratteristiche dei dati, il ché lo rende ideale per compiti di classificazione. Gli autoencoder sono utilizzati anche per la riduzione del rumore, la compressione dei dati e la generazione di nuovi dati.
Una variante dell'autoencoder è il Convolutional Autoencoder, in cui le operazioni di codifica e decodifica sono svolte da strati convoluzionali, il ché li rende appropriati per essere utilizzati con immagini.
Un utilizzo di un Convolutional Autoencoder è quello di [Cheng Ting-Yun 2020 - Identifying Strong Lenses with Unsupervised Machine Learning using Convolutional Autoencoder] in cui è stato utilizzato sulle immagini simulate per il telescopio spaziale Euclid. Questa tecnica, si è rivelata capace di identificare lensing gravitazionali di vario tipo.