

Introduzione e Obiettivi della Tesi (Slide 1-2: 30 secondi)

Buongiorno a tutti. Sono Martin Tomassi, e sono qui oggi per presentarvi la mia tesi di laurea: **"Valutazione Metodologica ed Applicativa di KAN, MLP, Random Forest e XGBoost con Tecniche di Ottimizzazione su differenti casi di studio"**.

La tesi è stata strutturata sulla base di due obiettivi principali:

1. **Studio Teorico**, dove ho Analizzato i fondamenti matematici e le architetture dei modelli, dedicando particolare attenzione alle **Kolmogorov-Arnold Networks (KAN)**.
2. **Analisi Metodologica ed Applicativa**, dove ho Valutato l'efficacia dei modelli in scenari reali, tramite:
3. **Addestramento e Valutazione**, dove ho provato i modelli su tre casi di studio diversi per tipologia di problema e natura dei dati.
4. **Studio di Ablazione**, dove ho analizzato l'impatto del **pruning** per misurare il trade-off tra la complessità del modello e le sue performance.

MLP e Backpropagation (Slide 3-4: 45 secondi)

Prima di esplorare le KAN, richiamo brevemente le **MLP**, che sono la base delle reti neurali tradizionali. Una MLP è composta da vari strati: uno di input, uno di output e vari strati nascosti. Il processo chiave in ogni neurone si svolge in due fasi: gli input subiscono una **somma ponderata** con i pesi ed il risultato viene poi fatto passare attraverso una **funzione di attivazione** non-lineare. L'apprendimento avviene tramite la **Backpropagation**: un algoritmo che propaga l'errore calcolato nello strato di output, dalla Loss, **all'indietro**. Sfruttando la **Chain Rule** e il **Gradient Descent**, i pesi e i bias vengono aggiornati in modo da minimizzare l'errore complessivo.

B-Spline e KART (Slide 5-6: 2 minuti e 30 secondi)

Il cuore matematico delle KAN risiede nelle **B-Spline**. Le B-Spline sono **funzioni polinomiali a tratti** utilizzate per definire le attivazioni nelle KAN in modo flessibile. Il loro utilizzo è cruciale perché risolve i principali limiti delle curve di Bézier. A differenza delle Bézier, che perdono controllo locale e diventano computazionalmente pesanti all'aumentare dei punti, le B-Spline sono:

- **Continue e differenziabili.**
- Offrono **controllo locale.**
- I coefficienti sono **addestrabili tramite backpropagation.**

Il teorema alla base dell'architettura delle KAN è il **Kolmogorov-Arnold Representation Theorem (KART)**. Questo teorema afferma che ogni funzione continua multivariata può essere espressa come una composizione di somme di funzioni continue univariate. Il limite,

tuttavia, è che il KART è solo **esistenziale**, non costruttivo: ci dice che la rappresentazione esiste, ma non ci fornisce un metodo esplicito per determinarla.

KAN (Slide 7-8: 2 minuti)

La vera innovazione delle **KAN** è come ridefinisce l'architettura neurale. Mentre le MLP hanno **pesi scalari** apprendibili sui nodi, le KAN sostituiscono questi pesi con **funzioni univariate parametriche ed apprendibili** poste direttamente sugli archi. Inoltre, le MLP alternano operazioni lineari a funzioni di attivazione fisse. Le KAN, invece, sostituiscono ogni prodotto peso-input con una di queste funzioni parametriche basate su B-Spline.

Questo approccio permette alle KAN di unire i vantaggi di entrambi i mondi:

- Dalle **Spline**: Efficacia a bassa dimensionalità e **controllo locale**.
- Dalle **MLP**: Mantengono la **Backpropagation**, la composizionalità e la scalabilità.

Il risultato più importante è che le KAN riescono a **superare la curse of dimensionality** che affligge le MLP. L'errore di approssimazione, infatti, dipende dalla risoluzione delle Spline, non dalla dimensionalità dell'input, portando a **scaling laws più favorevoli**.

Altri Modelli e Metodologia (Slide 9-12: 1 minuto e 15 secondi)

Durante lo studio, ho confrontato le KAN e MLP con altri modelli di riferimento nel machine learning:

- **XGBoost**: Un algoritmo di **ensemble sequenziale** in cui ogni albero corregge gli errori del precedente.
- **Random Forest**: Un algoritmo di **ensemble parallelo** che combina le previsioni di alberi addestrati su sottoinsiemi casuali di features e dati.
- **CNN**: Nel caso delle immagini, abbiamo applicato una modifica cruciale: i classificatori *fully-connected* standard sono stati sostituiti da una **KAN**, per un confronto diretto.

Per l'ottimizzazione degli iperparametri, ho scelto il **Random Search** per la sua **efficienza, semplicità di implementazione, scalabilità, parallelizzazione**.

L'intero processo è stato implementato utilizzando **PyTorch, pykan, scikit-learn e XGBoost**, sfruttando il **Cluster HPC dell'Università** per l'allenamento intensivo. Per la valutazione, ho standardizzato le metriche per regressione e classificazione, utilizzando gli **intervalli di confidenza bootstrap** per stimare l'incertezza dei risultati.

Casi di Studio (Slide 14-21: 2 minuti)

Passando ai risultati, la tesi si è concentrata su tre casi di studio, ma verranno presentati

solamente gli ultimi due.

Nel **Caso 2** mi sono occupato della **Classificazione dell'Inquinamento da PM2.5** in 6 classi AQI. Ho utilizzato un vasto dataset di serie storiche, raccogliendo dati su **453 città indiane per il periodo 2010-2023**, e il set di feature includeva 24 variabili, con *lag features* temporali per catturare la dinamica storica.

La fase di **Data Preparation** è stata fondamentale:

- Ho rimosso le colonne ridondanti o non significative alla classificazione.
- Ho gestito i Missing Values eliminando le colonne con più del 40% di dati mancanti ed i restanti sono stati sostituiti tramite forward-fill o con la media.
- Ho effettuato un ricampionamento a livello statale con media giornaliera per semplificare l'analisi.
- Ho rimosso gli Outlier, tramite Isolation Forest.

In questo scenario:

- In termini di accuratezza assoluta, il vincitore è stato **XGBoost**.
- Tuttavia, la **KAN** ha offerto il **miglior compromesso** tra performance e complessità, avendo un numero di parametri bassissimo, circa 7.7k, simile a MLP.
- Nello studio di ablazione, gli **Ensemble** (XGBoost e Random Forest) hanno mostrato una **straordinaria resilienza al pruning**, mantenendo prestazioni ottimali anche con compressioni elevate, mentre le Reti Neurali erano più fragili alla riduzione dei parametri.

Successivamente, nel **Caso 3**, ho affrontato la **Classificazione dell'Età da Immagini del Volto**, utilizzando il **Dataset UTKFace** per classificare l'età in 4 fasce predefinite.

La fase di **Data Preparation** è stata fondamentale:

1. Ho prima **bilanciato il dataset** campionando un numero quasi uguale di immagini per fascia d'età, per evitare che il modello privilegiasse le classi più rappresentate.
2. Successivamente, tutte le immagini sono state ridimensionate in forma quadrata e **normalizzate**.

Per quanto riguarda l'architettura, abbiamo confrontato **CNN+MLP** e **CNN+KAN**. Sebbene la **CNN+KAN** abbia avuto performance leggermente migliori (F1-Weighted 0.7843), la **CNN+MLP** è emersa come la soluzione più **efficiente**, offrendo un eccellente compromesso tra performance e complessità, grazie ai suoi 6 milioni di parametri, nettamente inferiori ai 16.4 milioni richiesti dalla CNN+KAN.

Conclusioni (Slide 22: 30 secondi)

In conclusione, il messaggio principale di questo studio è che **non esiste una soluzione**

universale. La scelta del modello ideale dipende sempre dalla natura del problema e dai vincoli operativi.

- **XGBoost** si è confermato il modello migliore per l'accuratezza sui dati tabulari.
- La **KAN** si è distinta per l'ottimo compromesso tra performance e complessità.
- Infine, è emersa la **straordinaria resilienza degli Ensemble** al pruning.

Le KAN, quindi, non sono un sostituto universale delle MLP, ma un'alternativa valida e promettente, specialmente dove è cruciale avere un modello **leggero e interpretabile**.

Grazie per l'attenzione.