3. REGULARIZAZIOTION FITTING

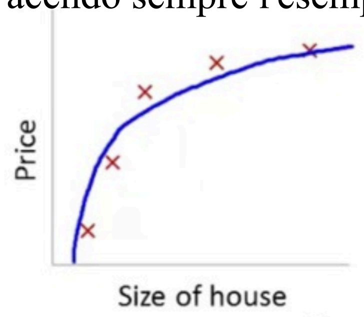
lunedì 28 ottobre 2024 11:16

Per affrontare il problema dell'overfitting, ovvero quando un modello si adatta troppo ai dati di addestramento perdendo capacità di generalizzazione, si possono utilizzare due approcci principali:

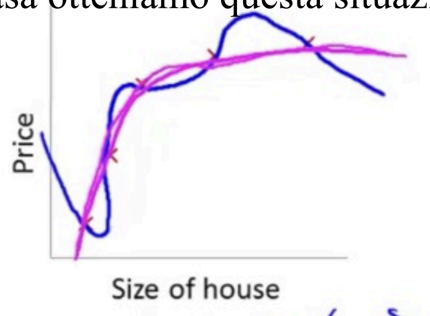
1. Ridurre il numero di caratteristiche (features)
Selezione manuale delle caratteristiche: consiste nel valutare manualmente quali caratteristiche sono rilevanti per il problema e rimuovere quelle meno significative. Questo approccio richiede una buona comprensione dei dati e del dominio del problema, ed è spesso efficace se il numero di caratteristiche è gestibile.
Algoritmi di selezione del modello: ci sono algoritmi automatici che possono aiutare a identificare le caratteristiche più rilevanti per il modello. Questi algoritmi analizzano le caratteristiche e selezionano solo quelle che forniscono un contributo significativo alla previsione, riducendo la complessità del modello e abbassando il rischio di overfitting. Alcuni esempi di tecniche di selezione sono l'eliminazione ricorsiva delle caratteristiche (RFE) o l'utilizzo di modelli basati su alberi decisionali, come l'algoritmo di random forest per stimare l'importanza di ciascuna caratteristica.
Ridurre il numero di caratteristiche può aiutare a ridurre l'overfitting perché un modello con meno caratteristiche è generalmente meno complesso e quindi meno incline ad adattarsi eccessivamente ai dati di addestramento.

2. Regolarizzazione
Mantenere tutte le caratteristiche, ma ridurre l'ampiezza dei parametri: la regolarizzazione è una tecnica che permette di mantenere tutte le caratteristiche nel modello, ma "penalizza" i valori eccessivi dei parametri, in modo da evitare che il modello si adatti troppo ai dati specifici.
Come funziona: la regolarizzazione aggiunge un termine di penalità alla funzione obiettivo del modello (solitamente l'errore di predizione), disincentivando parametri di grandi dimensioni. Esistono diverse forme di regolarizzazione, tra cui:
L2 (Ridge Regression): aggiunge una penalità proporzionale al quadrato dei parametri. Questo riduce gradualmente i valori dei parametri senza azzerarli completamente.
L1 (Lasso Regression): aggiunge una penalità proporzionale al valore assoluto dei parametri, il che porta alcuni parametri a diventare esattamente zero, selezionando implicitamente le caratteristiche più importanti.Quando è efficace: la regolarizzazione è particolarmente

utile quando si hanno molte caratteristiche, ciascuna delle quali contribuisce in parte alla previsione del target y. Invece di eliminare caratteristiche, la regolarizzazione permette di tenere tutte le informazioni, ma “attenuando” il peso dei parametri, rendendo il modello meno sensibile alle fluttuazioni specifiche dei dati di addestramento.In sintesi, ridurre il numero di caratteristiche e applicare la regolarizzazione sono due strategie fondamentali per ridurre l’overfitting, mantenendo un buon equilibrio tra accuratezza sui dati di addestramento e capacità di generalizzazione su nuovi dati.
Facendo sempre l'esempio della casa otteniamo questa situazione:



$\theta_0 + \theta_1 x + \theta_2 x^2$



$\theta_0 + \theta_1 x + \theta_2 x^2 + \cancel{\theta_3 x^3} + \cancel{\theta_4 x^4}$

$1000\,\theta_3 + 1000\,\theta_4 \qquad \theta_3 \approx 0 \qquad \theta_4 \approx 0$

Quando diciamo che vogliamo "ridurre il contributo dei termini di alto grado di un polinomio per rendere la curva più liscia," intendiamo limitare quanto i termini di alto grado influenzino la forma della curva. I termini di alto grado tendono a introdurre ondulazioni e fluttuazioni nel polinomio, il che può portare all’overfitting, ovvero quando il modello si adatta troppo ai dati di addestramento e perde capacità di generalizzazione.

Obiettivo della Regolarizzazione
La regolarizzazione serve a rendere più piccoli i valori dei parametri θ del modello. Questo porta ad avere una ipotesi più semplice e meno suscettibile all'overfitting.
In altre parole, avere valori piccoli per i parametri riduce la complessità del modello e permette di generalizzare meglio su nuovi dati.

Senza regolarizzazione abbiamo semplicemente la media dell'errore quadratico

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h(x^{(i)}) - y^{(i)} \right)^2$$

AGGIUNGIAMO QUINDI:

$$J(\theta) = \frac{1}{2m}\left(\sum_{i=1}^{m}\left(h\left(x^{(i)}\right)-y^{(i)}\right)^2 + \underline{\lambda\|\theta\|_2^2}\right)$$

Con la regolarizzazione viene aggiunto un termine di penalità $\lambda\|\theta\|_2^2$ che è proporzionale alla somma dei quadrati dei parametri θ

La norma L2 al quadrato, rappresentata da $\|\theta\|_2^2$ corrisponde alla somma dei quadrati dei parametri ECCETTO il termine di bias. Viene scritta come:

$$= \sum_{1}^{m} \theta_j^2$$

ORA LA NOSTRA C E' PROPRIO λ
DOVE RAPPRESENTA EFFETTIVAMENTE QUANTO SIA FORTE QUESTA PENALITA'

Se λ è grande, la penalità sarà forte, costringendo i valori dei parametri a ridursi maggiormente. Questo rende il modello più semplice, riducendo il rischio di overfitting ma aumentando il bias.
Se λ è piccolo, la penalità è debole, e quindi il modello può adattarsi di più ai dati di addestramento, ma con un rischio maggiore di overfitting.

$$J(\theta) = \frac{1}{2m}\left(\sum_{i=1}^{m}\left(h\left(x^{(i)}\right)-y^{(i)}\right)^2 + \lambda\sum_{j=1}^{n}\theta_j^2\right)$$

θ_0 NON INCLUSO

θ_0 RAPPRESENTA UNA COSTANTE, QUINDI SPOSTA SOLO IL MODELLO

NELLA REGRESSIONE LINEARE NOI SCEGLIAMO DI MINIMIZZARE. $J(\theta)$