

3. L1 & L2 NORM + slide

lunedì 28 ottobre 2024 17:28

L1 e L2 sono due modi per la regolarizzazione, due metodi per aggiungere un termine di penalità nella funzione di costo. Ecco una spiegazione dei concetti mostrati:

$$l_1 = \sum_{j=1}^m |\theta_j| \quad \text{L1}$$

Quando si usa la norma L1 come termine di penalità, si tende a ridurre alcuni parametri θ_j esattamente a zero, risultando in un modello più semplice. Questo effetto è utile per la selezione delle caratteristiche, poiché mantiene solo i parametri più importanti.

Tuttavia, la norma L1 non è derivabile (non ha una derivata) in zero, quindi la discesa del gradiente standard non può essere usata direttamente.

Cioè praticamente tende a eliminare i parametri per rendere il modello più semplice.

$$L2 \quad l^2 = \sum_{j=1}^m \theta_j^2$$

Questa norma è derivabile ovunque, quindi la discesa del gradiente può essere usata facilmente. L'uso della norma L2 tende a

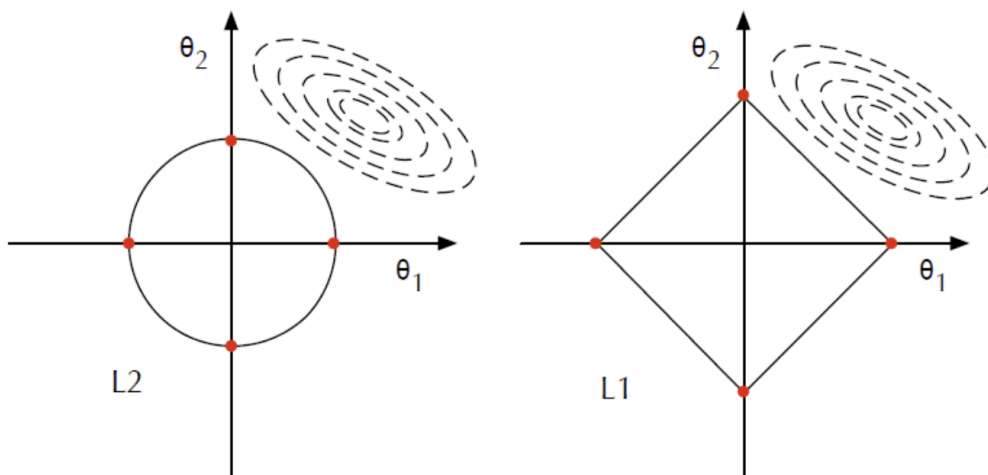
ridurre i valori dei parametri \(\sim\) senza azzerarli completamente, riducendo la complessità del modello mantenendo tutti i parametri

3. Problema di Ottimizzazione e Vincoli:

- Per minimizzare la funzione di costo (indicato come $\sum_{i=1}^m (\dots)^2$), possiamo usare una penalizzazione basata su una norma.
- **Vincolo con norma L2:** Si limita la somma dei quadrati dei parametri ($\|\theta\|_2^2 < t$).
- **Vincolo con norma L1:** Si limita la somma dei valori assoluti dei parametri ($\|\theta\|_1 < t$).

4. Esempio con Due Parametri:

- Quando l'analisi è limitata a due parametri (θ_1 e θ_2), le norme L2 e L1 assumono forme diverse:
 - Con la norma L2: il vincolo $\theta_1^2 + \theta_2^2 < t$ forma un cerchio.
 - Con la norma L1: il vincolo $|\theta_1| + |\theta_2| < t$ forma un rombo.
- Queste diverse forme dei vincoli implicano diversi effetti sui parametri. La norma L1 è più propensa a portare parametri esattamente a zero, mentre la norma L2 tende solo a ridurli.



Qui se facciamo la distanza tra il punto rosso e il disegnino tratteggiato notiamo dunque che in L2 non si azzerava completamente invece in L1 si. Quindi essenzialmente va ad eliminare i parametri.

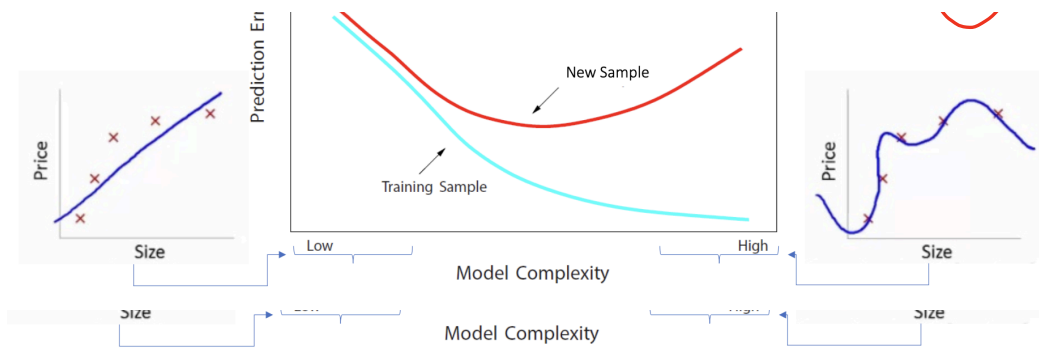
GRAFICI BIAS AND VARIANCE

Training set and test set w.r.t. complexity

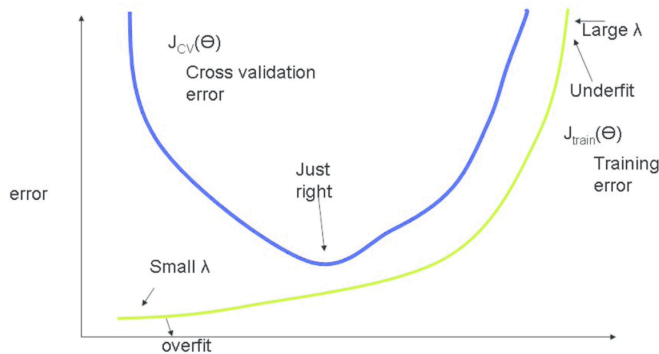


Let us recap the Bias variance behavior w.r.t. the complexity of the model





Training set and new samples w.r.t. regularization



②

Grafico in alto - Bias-Variance Tradeoff:

①

L'immagine centrale mostra come l'errore di previsione varia in funzione della complessità del modello.

Asse X (Model Complexity): Rappresenta la complessità del modello, da basso a sinistra (modelli semplici) a alto a destra (modelli complessi).

Asse Y (Prediction Error): Rappresenta l'errore di previsione.

Quando il modello è troppo semplice, si ha un alto bias e bassa varianza (il modello non riesce

a catturare le tendenze dei dati). Questo si verifica a sinistra del grafico.

All'aumentare della complessità del modello, l'errore si riduce fino a un punto ottimale (centro del grafico), dove c'è un buon equilibrio tra bias e varianza.

Se la complessità continua a crescere, l'errore sui nuovi campioni aumenta, poiché il modello tende ad adattarsi troppo ai dati di addestramento, causando basso bias ma alta varianza (overfitting).

Le immagini a sinistra e destra del grafico rappresentano visivamente un modello con basso bias (alto adattamento ai dati) e un modello con alto bias (basso adattamento ai dati).

Grafico in basso - Effetto della Regularizzazione:

Questo grafico descrive come la regularizzazione (parametro λ) influisce sull'errore di training e sull'errore di cross-validation.

Asse X (λ): Indica il valore della regularizzazione, da basso (sinistra) a alto (destra).

Asse Y (error): Rappresenta l'errore, sia per il training che per la cross-validation.

La curva verde (J_{train}) rappresenta l'errore sul set di addestramento, mentre la curva blu (J_{Θ}) rappresenta l'errore di cross-validation.

λ piccolo: Porta a un overfitting, con un errore di training molto basso ma un errore di cross-validation alto, poiché il modello si adatta troppo ai dati di training.

λ grande: Porta a un underfitting, dove l'errore sia di training che di cross-validation è alto, poiché la regolarizzazione troppo forte limita la capacità del modello di adattarsi ai dati.

C'è un valore ottimale di λ in cui l'errore di cross-validation è minimo, indicato come "Just right", dove il modello ha un buon compromesso tra complessità e capacità di generalizzazione.