

Machine Learning - Challenge 2

Abdula Kalus

Svolgimento challenge 2, corso di Intelligenza Artificiale, Università degli Studi di Trieste

Abstract

In questa challenge si vogliono mettere in pratica gli argomenti studiati a lezione riguardo alle SVM ed i Kernel applicandoli a concetti già visti come regressione lineare e PCA. L'obiettivo è quello di verificare l'effettiva efficacia dei kernel nell'analizzare e interpretare dati non linearmente separabili.

Ridge Regression

In questo primo caso di studio si vuole studiare l'andamento di una funzione non lineare

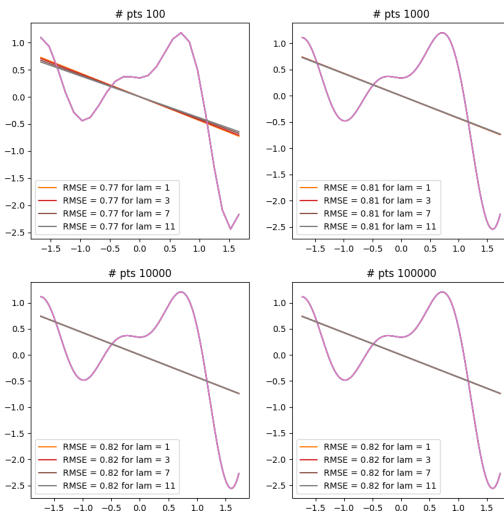


Fig. 1. Ridge regression con diversi α

Come possiamo vedere dalla figura 1 per quanto il numero di punti aumenti e variamo l'iperparametro α per la regressione con regolarizzazione ridge, non otteniamo mai un buon modello, per quanto l'errore quadratico medio sembri buono.

Invece usando KRR (Kernel Ridge Regression) riusciamo ad ottenere degli ottimi risultati, come si può vedere dalla figura 2, nella quale sono presenti diverse esecuzioni del metodo con diversi parametri e per diverse quantità di osservazioni. Ciò che salta all'occhio è che il kernel lineare underfitta i dati e non è sufficientemente espressivo un modello con questo tipo di kernel, invece per quanto riguarda un modello con un kernel gaussiano, possiamo vedere che in base a come vengono settati i parametri e al numero di dati che abbiamo a disposizione il modello può diventare molto buono, maggiore è il numero di dati a disposizione migliore può essere l'apprendimento del modello. ad esempio osservando

la prima colonna possiamo osservare che già con 100 punti si riesce a intuire la funzione vera, ed all'aumentare dei punti il modello tende ad aumentare.

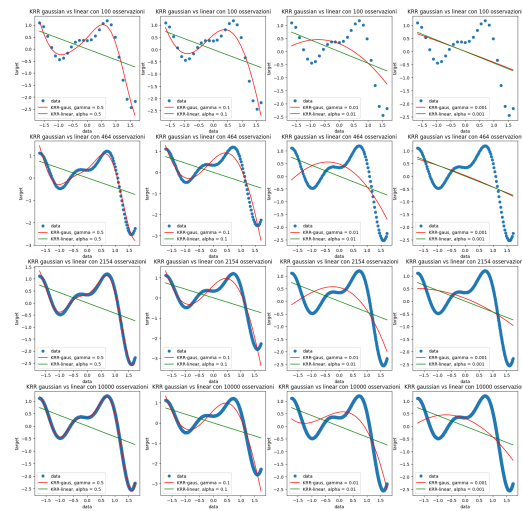


Fig. 2. KRR esempi multipli

Infatti utilizzando GridSearch su un dataset composto da 2000 osservazioni otteniamo che i migliori iperparametri per KRR con kernel gaussiano sono $\alpha = 0.001$, $\gamma = 1.73$ con un $R^2 = 0.942$

PCA

In questa sezione consideriamo un esempio di classificazione dove non è possibile tracciare in modo semplice una retta per separare i punti, infatti ci sono diversi problemi con il dataset in figura 3, il primo è che i campioni di ciascuna classe non possono essere separati linearmente, ovvero nessun piano può dividere i campioni dell'anello interno dall'anello esterno.

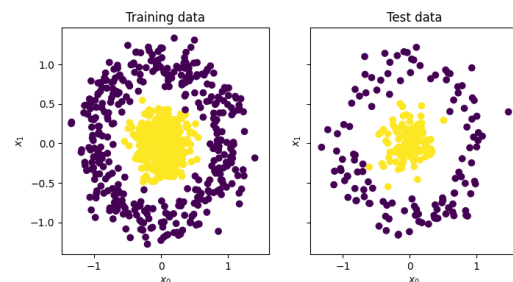


Fig. 3. dataset circles

Infatti se tracciamo la proiezione su entrambe le componenti principali, vediamo nella figura ?? che i dati sembrano uguali, o isotropi, nel senso che hanno lo stesso valore

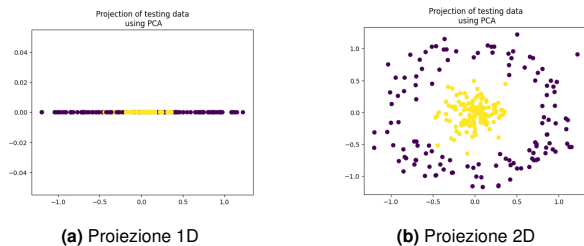


Fig. 4. Proiezioni

quando misurati in un sistema di coordinate diverso. Questo perché i dati sono simmetrici e la varianza è uguale su ciascun lato. Di conseguenza, la PCA non ci aiuterà a classificare i dati.

Invece applicando PCA dopo una trasformazione polinomiale dei dati, ovvero $\phi(x) = [x_1, x_2, (x_1)^2 + (x_2)^2]$ dove $x = [x_1, x_2]$, e proiettando su un piano formato dalle prime due componenti principali (quelle con gli autovalori assoluti più grandi), vediamo che il set di dati è linearmente separabile ad eccezione di un paio di punti.

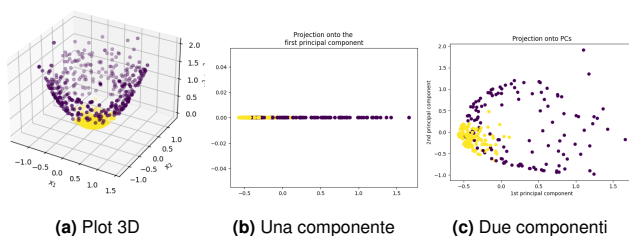


Fig. 5. PCA con trasformazione polinomiale

ed infine utilizzando kernel PCA con un kernel Gaussiano possiamo osservare i seguenti risultati

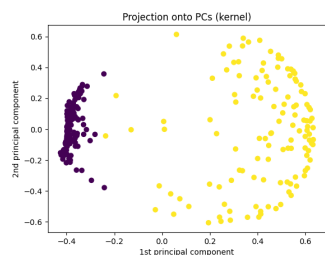


Fig. 6. Kernel PCA

dove è presente una netta separazione dei punti.

Nella tabella 1 possiamo osservare il valore dell'accuracy con il metodo SVM per tre diversi dataset, ovvero quelli nominati precedentemente, il primo il cui usa PCA senza nessun tipo di trasformazione, il secondo che usa PCA dopo aver aggiunto una features in più al datasete l'ultimo che utilizza PCA con un kernel gaussiano.

Tipo kernel PCA	accuracy
nessuno	0.496
trasformazione	0.996
gaussiano	0.992

Table 1. Precision measures

Dataset con molto rumore

Nel seguente esempio applichiamo le tecniche precedenti su un dataset molto rumoroso, dove i punti sono molto mischiati tra di loro, come possiamo vedere nel plot successivo ??

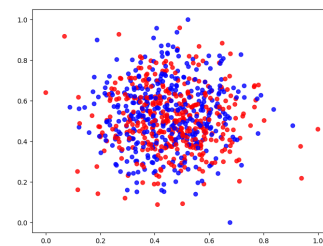
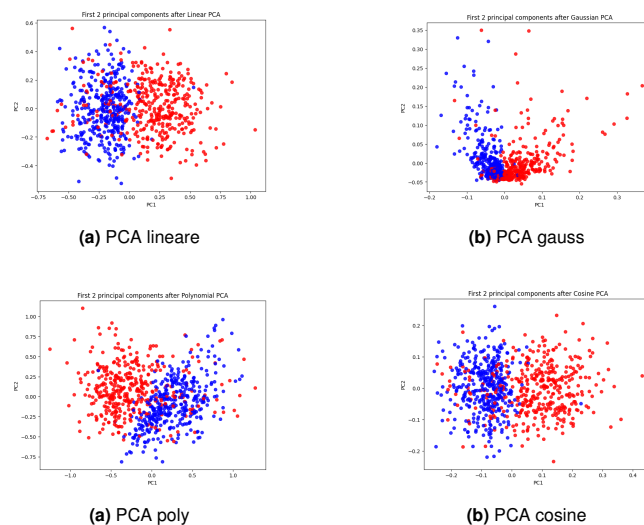


Fig. 7. dataset

Applicando PCA con diversi meotodi di kernel possiamo osservare i segeunti risultati ??



Si può osservare che pur avendo un dataset "ingestibile" nel quale non sembrano essere presenti delle correlazioni tra le features, in realtà andando in uno spazio dimensionale più alto esistono delle correlazioni che ci permettono di classificare, come is può vedere nella tabella 2 applicando PCA con diversi kernel e successivamente un classificatore SVM otteniamo i seguenti riusltati

Tipo kernel PCA	accuracy
lineare	0.904
gaussiano	0.952
polinomiale	0.948
coseno	0.94

Table 2. Precision measures