

Risorse

Il codice utilizzato, insieme al file .tex di questo documento, possono essere trovati nella seguente repository github:
https://github.com/LazyLagrangian/data_science.

Esercizio 1 - Fasi del Modello di Ising con Regressione Logistica

L'obiettivo dell'esercizio è quello di allenare un regressore logistico su un dataset di modelli di Ising 2-D, classificati in ordinati e disordinati.

I dati forniti consistono in 160000 sistemi composti da 40×40 valori, disposti dalle temperature più basse alle temperature più alte (e quindi da più ordinati a più disordinati). Questi sono stati divisi in tre batch: i primi 70000 hanno formato il batch contenente i sistemi ordinati, i successivi 30000 il batch dei sistemi critici (al limite tra ordinati e disordinati), e i dati rimanenti il batch dei sistemi disordinati.

La divisione in batch non è da confondere con la classificazione dei sistemi: infatti, nonostante ciascun membro del batch ordinato sia classificato come ordinato (e analogamente per il batch disordinato), non vi è una classificazione a parte per i dati critici, che invece risultano essere ordinati o disordinati.

La figura 1 mostra sistemi tipici appartenenti a ciascun batch, la figura 2 fornisce una panoramica generale del livello di ordine relativo a ciascuno di essi.

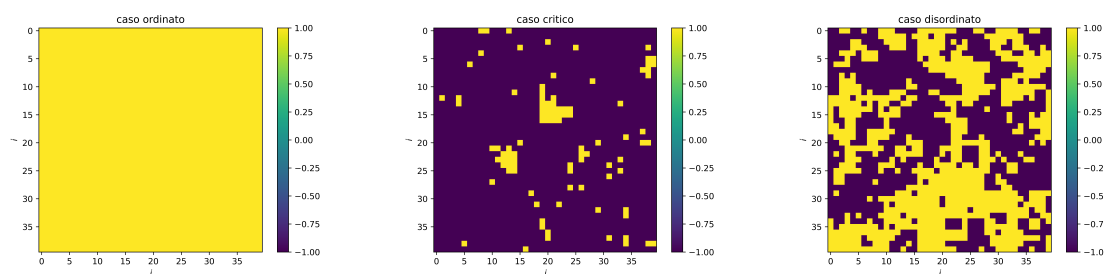


Figura 1: Tre esempi rappresentativi dei batch, ordinato, critico e disordinato. Ogni cella del piano x - y rappresenta uno dei 40×40 valori che caratterizzano un sistema.

Per il training sono stati utilizzati solamente i dati appartenenti al batch ordinato e disordinato, che sono stati ulteriormente suddivisi in batch di training e testing.

Si è poi allenato il regressore logistico, utilizzando tre metodi regolarizzati (secondo regolarizzazione L_2) proposti dalla libreria sklearn: lbfgs, liblinear e newton-cg. La performance R^2 dei vari metodi in funzione del parametro di regolarizzazione, è presentata in figura 3.

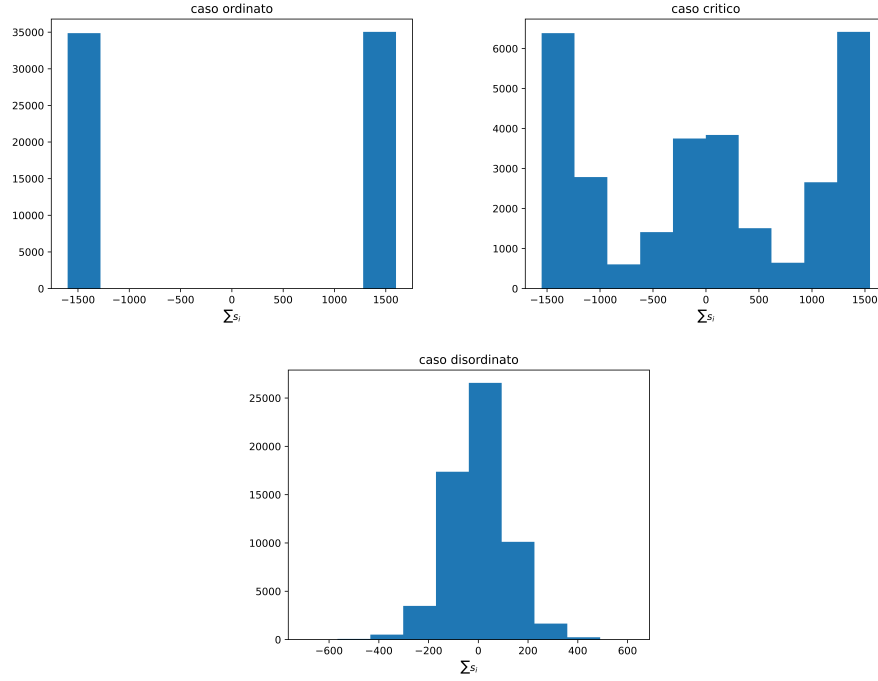


Figura 2: Per ciascun batch, la figura mostra un istogramma della somma dei valori che compongono i sistemi. Tale somma ci fornisce una misura del livello di ordine di un sistema.

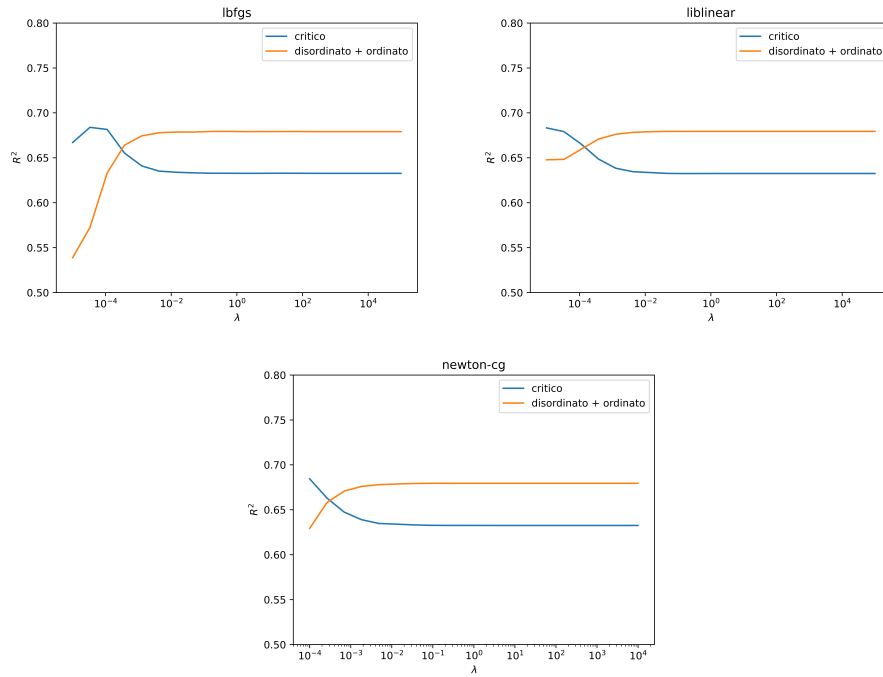


Figura 3: Sull'asse delle x il parametro di regolarizzazione λ , sull'asse delle y il coefficiente di determinazione R^2 , valutato sia sul batch di test degli ordinati e disordinati, che sul batch dei critici. I tre grafici sono relativi ciascuno ad uno dei metodi di regressione.