

**Esercizi per il corso di Data Science - Laurea in Scienza dei Materiali**

PROF. D. DI SANTE, DR. A. CONSIGLIO    **4° Foglio, Regressione Logistica e Modelli**  
 SEMESTRE INVERNALE 2024/2025    06/11/2024

**Esercizio 1 - Fasi del Modello di Ising con Regressione Logistica**

L'obiettivo di questo esercizio è utilizzare la regressione logistica per classificare gli stati del modello di Ising 2D in base alla loro fase. L'energia (Hamiltoniana) del modello è data da:

$$H = -J \sum_{\langle ij \rangle} S_i S_j, \quad S_{i,j} \in \{-1, +1\}$$

dove gli indici del sito reticolare  $i, j$  si estendono su tutti i primi vicini di un reticolo quadrato 2D.  $J$  è una scala arbitraria di energia di interazione.

Onsager ha dimostrato che questo modello presenta una transizione di fase in funzione della temperatura nel limite termodinamico: da un ferromagnete ordinato con tutti gli spin allineati (per temperature basse) ad una fase disordinata al di sopra di una temperatura critica  $T_C/J \sim 2.26$ .

Dato uno stato di Ising, vorremmo classificare se esso appartiene alla fase ordinata o disordinata, senza alcuna informazione aggiuntiva oltre alla configurazione di spin stessa. Questo problema categorico di apprendimento automatico è adatto per la regressione logistica.

**(a)** Consideriamo il modello di Ising 2D su un reticolo quadrato  $40 \times 40$  ed un gran numero di stati ( $10^4$  stati) per ciascuna delle 16 temperature  $T$  prese da un insieme predefinito  $T \in [0.25, 0.5, \dots, 4.0]$ . Gli stati sono stati generati da un algoritmo Monte Carlo e, utilizzando il criterio di Onsager, possiamo assegnare un'etichetta a ciascuno stato in base alla sua fase: 0 se lo stato è disordinato e 1 se è ordinato. L'obiettivo è prevedere la fase di appartenenza di un campione data la configurazione degli spin.

È noto che, in prossimità della temperatura critica  $T_C$ , la lunghezza di correlazione ferromagnetica diverge il che, tra l'altro, porta ad un rallentamento critico dell'algoritmo per generare gli stati. Pertanto, ci si aspetta che l'identificazione delle fasi sia più difficile nella regione critica. Tenendo questo a mente, abbiamo considerato i seguenti tre tipi di stati: ordinato ( $T/J < 2.0$ , i primi 70000 stati), critico ( $2.0 \leq T/J \leq 2.5$ , stati con indice tra 70001 e 100000) e disordinato ( $T/J > 2.5$ , stati oltre l'indice 100000). Gli stati e le etichette sono contenuti nei relativi files pkl (pickles) che possono essere scaricati da Virtuale e letti attraverso il codice che trovate di sotto in figura.

Si visualizzino degli stati tipici in ciascuna delle tre regioni.

**(b)** Si utilizzino insieme solo stati ordinati e disordinati per addestrare il regressore logistico e, una volta completata la procedura di addestramento, si valutino le prestazioni del modello di classificazione su stati di test ordinati, disordinati e critici.

Come regressore logistico si usi la funzione `LogisticRegression` di `sklearn`. Si visiti la pagina web relativa per capire come usare la funzione. Si valutino le prestazioni del modello

in funzione della regolarizzazione, e dei diversi solver (ad esempio liblinear, lbfgs, newton-cg).

```
import pickle

##### LOAD DATA
# The data consists of 16*10000 samples taken in  $T=np.arange(0.25,4.0001,0.25)$ :
data_file_name = "Ising2DFM_reSample_L40_T=All.pkl"
# The labels are obtained from the following file:
label_file_name = "Ising2DFM_reSample_L40_T=All_labels.pkl"

#DATA
with open(data_file_name, 'rb') as f:
    data = pickle.load(f)
data = np.unpackbits(data).reshape(-1, 1600) # Decompress array and reshape for
↳ convenience
data=data.astype('int')
data[np.where(data==0)]=-1 # map 0 state to -1 (Ising variable can take values
↳ +/ -1)

#LABELS (convention is 1 for ordered states and 0 for disordered states)
with open(label_file_name, 'rb') as f:
    labels = pickle.load(f)
```