CORRIERE DELLA SERA

UNIVERSITÀ DI SIENA 1240

DIISM

CINECA

# DISCLAIMER

## ULTIME AVVERTENZE PRIMA DELLA RIVOLUZIONE

## BIOMIND

### INTELLIGENZA ARTIFICIALE PER L'INNOVAZIONE IN FARMACIA E BIOINFORMATICA

KNOWLEDGE PARTNER

FONDAZIONE Mondo digitale

Diritto alla conoscenza

# INTRODUZIONE - TIPO DI PROBLEMA

- Task di **regressione** con approccio MLP

- **Machine Learning** Task: **predire** un valore numerico continuo, date delle **features** in ingresso che descrivono le proteine

- Finalità applicative: predire l'**affinità di legame** tra la proteina *PID_170* e vari composti molecolari di riferimento, presi dal **dataset** fornito

# CONCETTI CHIAVE - ARCHITETTURA DELLA RETE

- IDEA: **selezione dell'architettura** più performante per gli strati nascosti dell'MLP tra varie configurazioni

- Numero features di **ingresso**: 14
- Numero features di **uscita**: 1 (regressione univariata)
- Best case (**hidden layers**): [100-200-130-50]

- Elementi principali: funzione di attivazione ReLU, rete **deep**

# COME FUNZIONA - PROCEDURA DI ADDESTRAMENTO

- **Preprocessing** dei dati: **normalizzazione** di ogni feature nell'intervallo [0,1] tramite *MinMaxScaler*

- Processo di **addestramento**: algoritmo di backpropagation
  - Optimizer: ADAM
  - Learning rate = 0.001
  - Max epochs = 1000
  - **Shuffle** del dataset tra le iterazioni

# VANTAGGI E LIMITI

- ✅ Vantaggi:
  - Flessibilità: **adattamento** dell'algoritmo ai dati specifici, i quali sono oggetto di studio
  - Robustezza: procedura di addestramento che garantisce **robustezza statistica**
  - Accuratezza: **misure di accuratezza** sui risultati finali, usando valori di riferimento (92.8% accuracy sul Test Set)
  - Velocità rispetto ad esperimenti in laboratorio

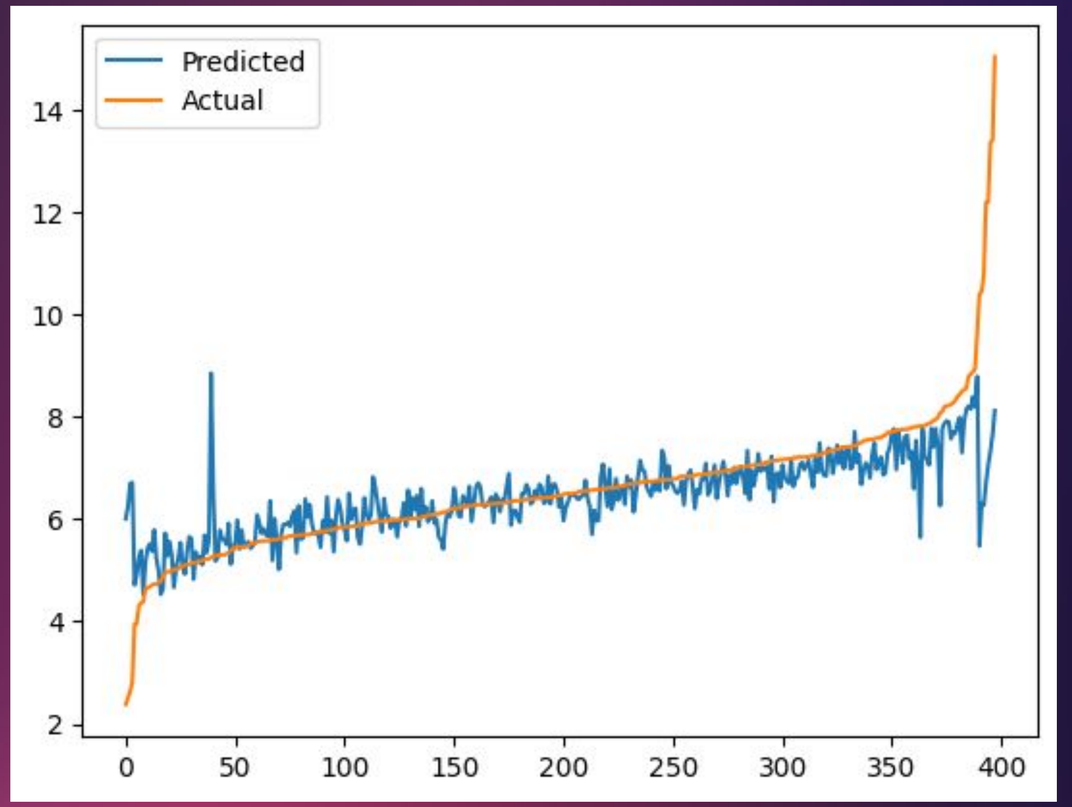
- ❌ Limiti:
  - Analisi basata su dati numerici e **non strutturati**
  - **Difficile interpretrabilità**, intrinsecamente data dalla natura delle layer"hidden"

# BEST RESULTS FOR AFFINITY (LOWER IS BETTER)

| | | | |
|---|---|---|---|
| 0 | CID_00083 | PID_170 | 4.857362 |
| 1 | CID_00232 | PID_170 | 5.298012 |
| 10 | CID_01679 | PID_170 | 5.785010 |

Plot della procedura di training (x-axis: epochs)

Le molecole sopra riportate mostrano la migliore affinità rispetto alla proteina di riferimento.



CORRIERE DELLA SERA

UNIVERSITÀ DI SIENA 1240

DIISM

CINECA

# GRAZIE DELL'ATTENZIONE!

# DISCLAIMER

ULTIME AVVERTENZE PRIMA DELLA RIVOLUZIONE

## BIOMIND

INTELLIGENZA ARTIFICIALE PER L'INNOVAZIONE IN FARMACIA E BIOINFORMATICA

KNOWLEDGE PARTNER

FONDAZIONE Mondo digitale

Diritto alla conoscenza