

# UDE e ClusterGAN per Clustering non-supervisionato nel Federated Learning

Lorenzo Sani

Università degli Studi di Bologna

July 21, 2021

# Federated Learning (FL) in breve

Caratteristiche generali: [1] [3]

- evitare la necessità di costituire grandi dataset "aggregati"
- sfruttare le nuove tecniche di deep learning (DL)
- definire la migliore procedura di aggregazione

Vantaggi:

- quasi nessuna limitazione sui dati dovuta alla privacy
- migliore sfruttamento delle risorse di calcolo (sale server)
- il framework Flower è semplice e "modellabile" [2]

Svantaggi:

- i principali algoritmi di aggregazione si adattano bene solo alle reti neurali

# Unsupervised Deep Embedding (UDE) for Clustering

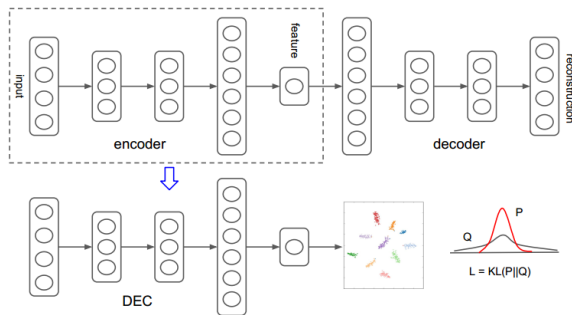
Obiettivi:[5]

- apprendere una rappresentazione dei dati (riduzione della dimensionalità)
- assegnare ai dati etichette dei cluster usando Deep Neural Network (DNN)

Uso:

- addestrare un autoencoder per la riduzione della dimensionalità
- estrarre l'encoder
- costituire in cima all'encoder un particolare Layer di Clustering

# Layer di Clustering



- distribuzione di probabilità ausiliaria basate sui centroidi dei cluster assegnati
- minimizza la divergenza di Kullback-Leibler (KL) della distribuzione stessa
- aggiornare dopo un certo numero di epoche la distribuzione

# ClusterGAN

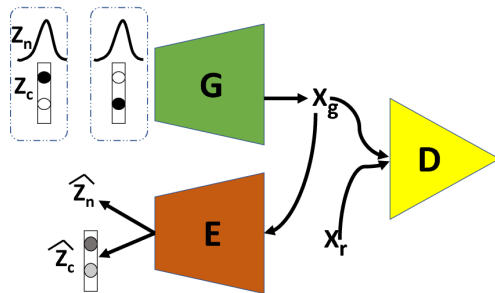
## Generative Adversarial Networks (GANs)

- due reti vengono addestrate in competizione l'una con l'altra
- si definisce generatore la rete che cerca di creare "dati artificiali" verosimili
- si definisce discriminatore la rete che cerca di scoprire se il set di dati è "reale" o "artificiale"

## Premesse alla ClusterGAN[4]

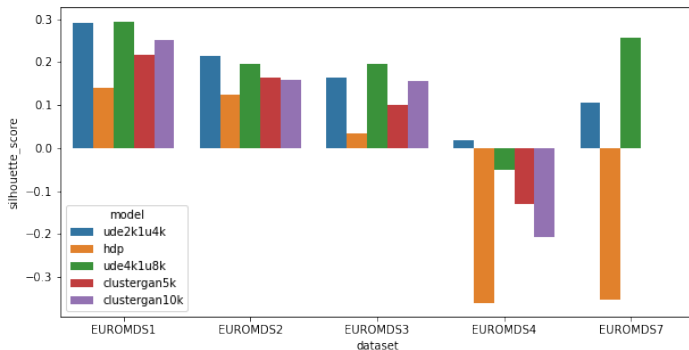
- le GAN hanno ottenuto un ampio successo in molti problemi non-supervisionati
- sfruttare l'abilità delle GAN di fare back-projection sullo spazio latente
- costruire una rete GAN capace di fare clustering nello spazio latente

# ClusterGAN

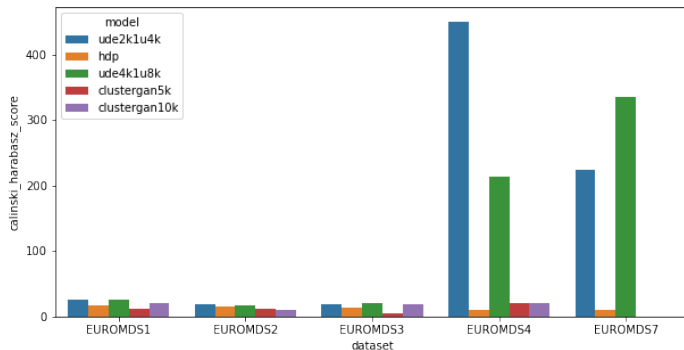


- campionare casualmente variabili latenti da un mix tra vettori "one-hot"  $z_c$  e valori continui  $z_n$
- "dare in pasto" i valori campionati alla GAN e i dati reali  $x_r$  al discriminatore
- accoppiare un encoder alla GAN che proietti i dati generati  $x_g$  sullo spazio latente
- utilizzare una loss specifica per il clustering

# Silhouette Score



# Calinski Harabasz Score





# Bibliografia



M. Aledhari, R. Razzak, R. M. Parizi, and F. Saeed.

Federated learning: A survey on enabling technologies, protocols, and applications.

*IEEE Access*, 8:140699–140725, 2020.



Daniel J. Beutel, Taner Topal, Akhil Mathur, Xinchu Qiu, Titouan Parcollet, Pedro P. B. de Gusmão, and Nicholas D. Lane.

Flower: A friendly federated learning research framework, 2021.



H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas.

Communication-efficient learning of deep networks from decentralized data, 2017.



Sudipto Mukherjee, Himanshu Asnani, Eugene Lin, and Sreeram Kannan.

Clustergan : Latent space clustering in generative adversarial networks.

*CoRR*, abs/1809.03627, 2018.



Junyuan Xie, Ross B. Girshick, and Ali Farhadi.

Unsupervised deep embedding for clustering analysis.

*CoRR*, abs/1511.06335, 2015.

Fine,  
Grazie per l'attenzione