ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

XXXIII Cycle

Machine Learning and Information Theory

Giulio Franzese

Supervisor: Prof. Monica Visintin

Research context and motivation

- Bidirectional interaction between Machine Learning and Information Theory: use Information Theory concepts to design new Machine Learning algorithms (or extend existing ones) and use known Machine Learning techniques (such as Neural Networks) to Information Theoretic problems

Addressed research questions/problems

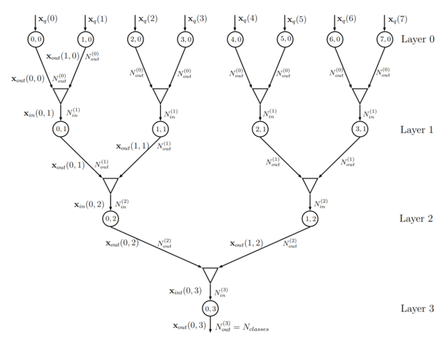
- Study of ensemble of Deep Information Network classifiers
- Feasibility of Mutual Information Neural Estimation scheme as a tool for telecommunication receivers
- Analysis of the posterior collapse problem and (mutual information vanishing) between input and and latent code in Variational Autoencoders

Novel contributions

- An extension of the previously proposed DIN algorithm has been proposed
- Mutual Information Neural Estimation has been shown to be a valid and principled alternative to derive a receiver for nonlinear telecommunications channel with memory
- A new information maximization cost function has been derived for variational autoencoding

Probabilistic Ensamble of Deep Information Networks (DIN)

- An information theoretic classification algorithm has been previously proposed (DIN)
- An ensembling method was designed to combine multiple DINs
- We showed the application of this algorithm to several UCI benchmarks

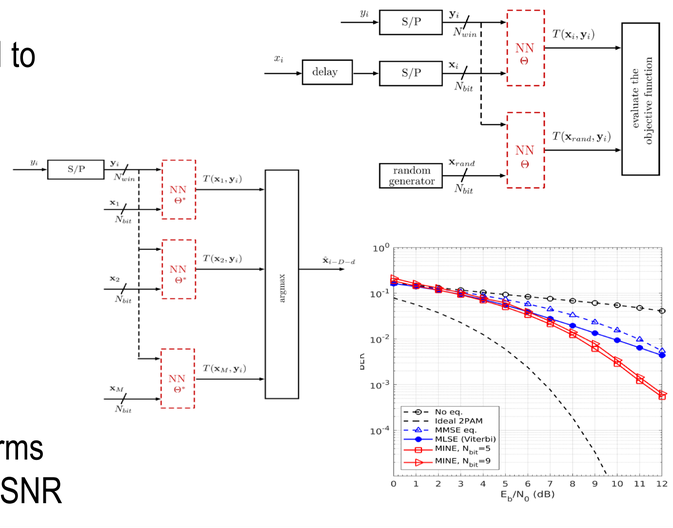


Mutual Information Neural Estimation Receiver

- Mutual Information (MI) can be written as a KL divergence $I(X;Y) = D_{KL}(p_{X,Y}(x,y)\|p_X(x)p_Y(y)) = \int\int p_{X,Y}(x,y)\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\,dx\,dy$
- When pdfs are not analytically known it is difficult to estimated MI (especially when the dimensionality is high)
- Thanks to functional convex duality we can express MI using the Donsker-Vahardan representation $D_{KL}(p_{X,Y}(x,y)\|p_X(x)p_Y(y)) \geq \sup_{T\in\mathcal{F}} \mathbb{E}_{p_{X,Y}(x,y)}\{T(X,Y)\} - \log\left(\mathbb{E}_{p_X(x)p_Y(y)}\left\{e^{T(X,Y)}\right\}\right)$
- The idea is to choose T to be a neural network and optimize the following cost function parametrized by θ $f(X,Y,\Theta) = \mathbb{E}_{p_{X,Y}(x,y)}\{T(X,Y;\Theta)\} - \log\left(\mathbb{E}_{p_X(x)p_Y(y)}\left\{e^{T(X,Y;\Theta)}\right\}\right)$
- We use this formalism to build a neural network that estimates the MI between a window of ISI affected received samples (x) nonlinearly distorted and the true transmitted symbols (y). We tested the method on a 2PAM modulated signal.
- As a byproduct, a receiver is obtained: the output of the optimal neural network is

$$T(X,Y;\Theta^*) = \log\left(\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\right) = \log\left(\frac{p_{X|Y}(x|y)}{p_X(x)}\right)$$

- During training the network is optimized to distinguish between true and randomly generated symbols
- During testing multiple copies of the network are used in parallel, and the symbol with highest metric is declared as the received one
- Results show that the solution outperforms classical alternatives especially at high SNR



Alternative cost function for Kernel Implicit VAE Information Maximization

- Variational Autoencoders are powerful unsupervised learning methods

$$X \longrightarrow \boxed{\mathscr{E}} \longrightarrow Z \longrightarrow \boxed{\mathscr{D}} \longrightarrow \hat{X}$$

- Input X is encoded into latent variable Z and decoded as $\hat{X}$
- Classical VAE are parametric (encoder and decoder generate parameters of Gaussian distributions)
- The cost functional to be optimized is

$$\mathscr{L}\{q,r\} = \int p_X(x)q(z|x)\log(r(x|z))dxdz - \int p_X(x)q(z|x)\log\frac{q(z|x)}{p_Z(z)}dzdx$$

- This functional can achive optimum also when MI between X and Z is 0 (posterior collapse problem)
- We propose a different cost functional that at the optimum maximizes MI between X and Z

$$\mathscr{B}\{q,r\} = \int p_X(x)q(z|x)\log(r(x|z))dxdz - \int p_Z(z)\log\frac{p_Z(z)}{q(z)}dz$$

- Our implementation is based on Kernel Density estimation
- Generation experiments: sample randomly Z and generate outputs

- Interpolation experiments: take leftmost image, encode it, take rightmost image and encode it, generate images by using latent variables Z obtained by linearly interpolating between the two initial latent representations

Adopted methodologies

- Personal Matlab Framework for Neural Networks based research problems
- Object oriented Matlab implementation of Probabilistic Ensemble of DIN

Future work

- Further investigate Information Theoretic techniques to be applied to Machine Learning
- Look for opportunities of application of Machine Learning algorithms to Information Theory problems

List of attended classes

02LWHRV Communication
01QRQRV Compressed sensing: theory and applications
01SHMRV Entrepreneurial Finance
08IXTRV Project management
01RISRV Public speaking
01SYBRV Research integrity
01SWQRV Responsible research and innovation, the impact on social challenges
02RHORV The new Internet Society: entering the black-box of digital innovations

Submitted and published works

- Mine Receivers, G.Franzese, M. Visintin, (submitted IEEE Electronic Letters)
- Probabilistic Ensemble of Deep Information Networks, G.Franzese, M. Visintin, (submitted MDPI Entropy)
- Alternative Cost Function for Kernel Implicit Variational Autoencoding based on Information Maximization , G.Franzese, M. Visintin, WIRN 2019 Vietri sul Mare

POLITECNICO DI TORINO

PhD program in
Electrical, Electronics and Communications Engineering