

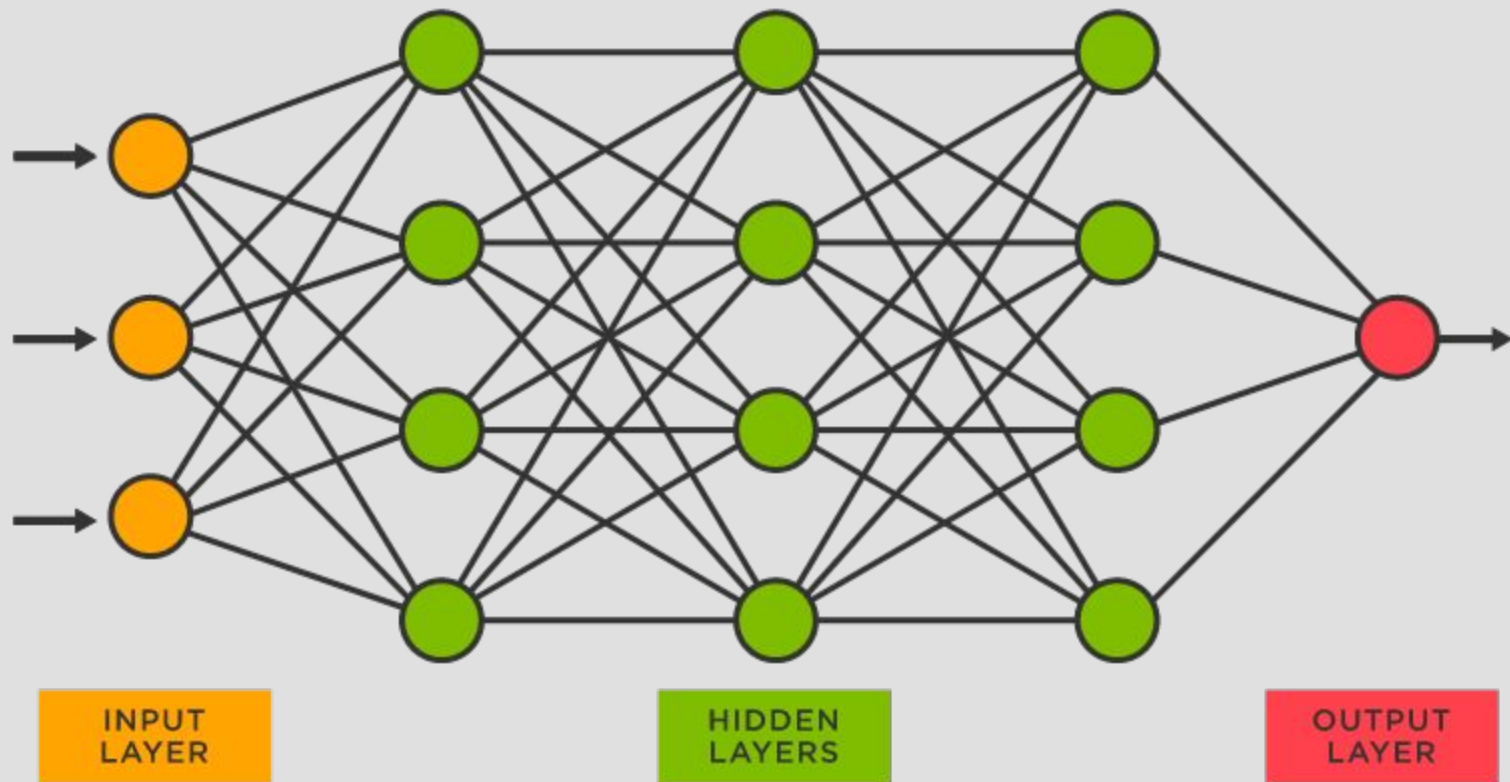
Comparação de Modelos de Representação Latente

...

José Geraldo Fernandes

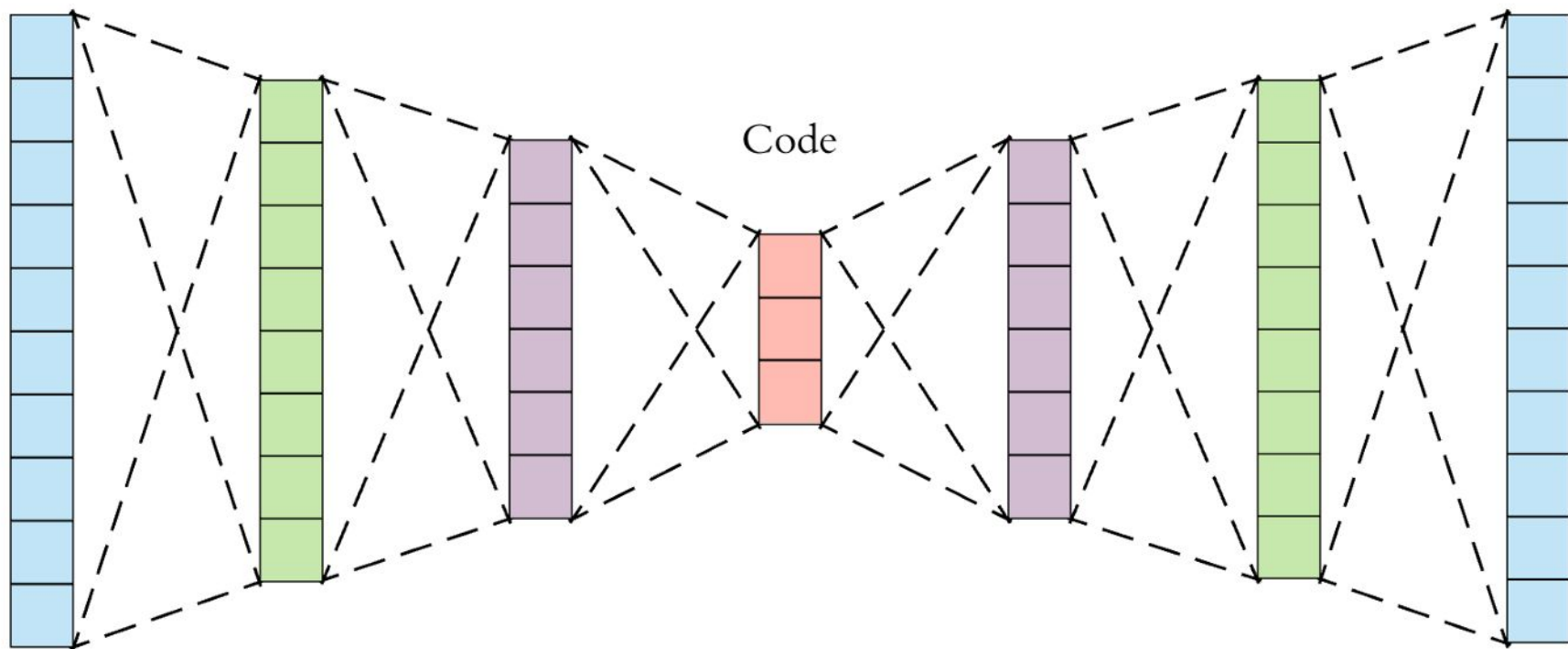
Aprendizado de Representação

- MLP
- Autoencoder



Input

Output



Code

Encoder

Decoder

$$\mathbf{W} = (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I}_p)^{-1} \mathbf{H}^T \mathbf{y}$$

$$\hat{\mathbf{y}} = S(\mathbf{H}\mathbf{W})$$

Motivação

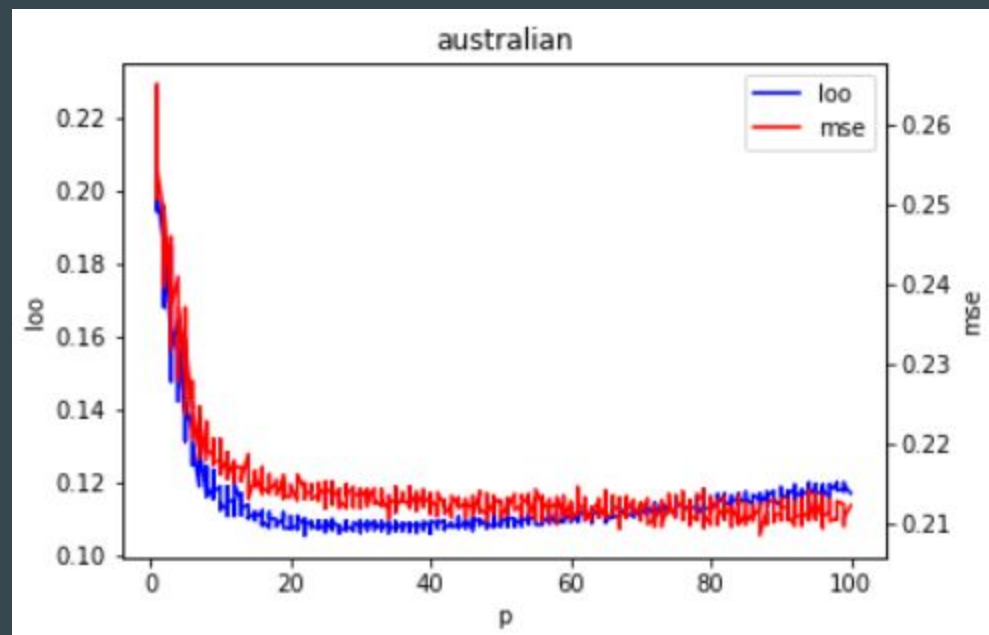
$$\mathbf{A} = (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I}_p)^{-1}$$

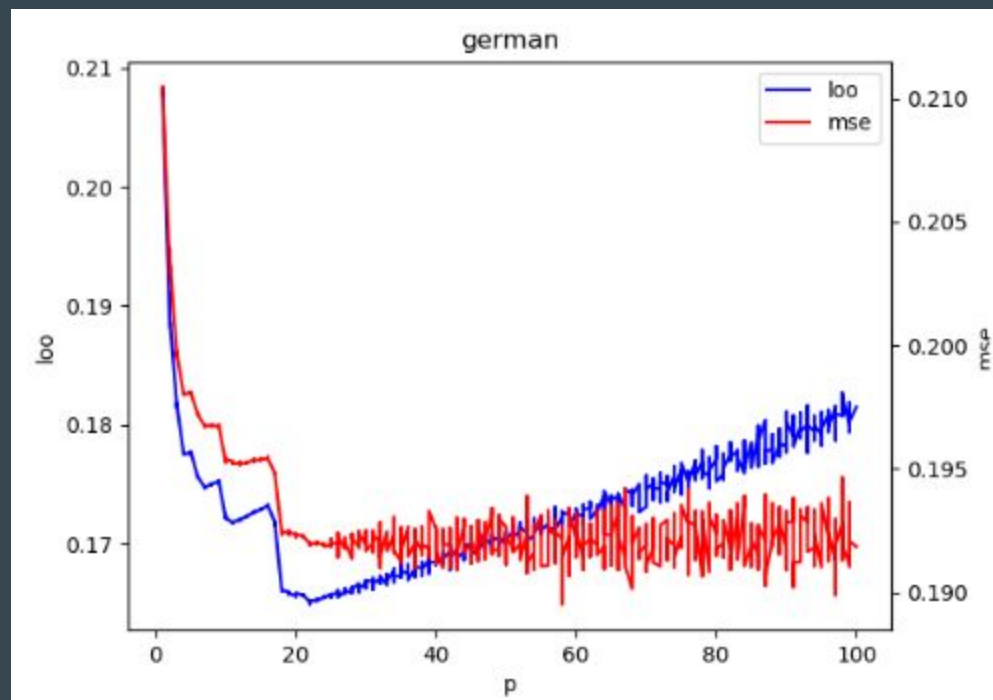
$$\mathbf{P} = \mathbf{I}_N - \mathbf{H} \mathbf{A} \mathbf{H}^T$$

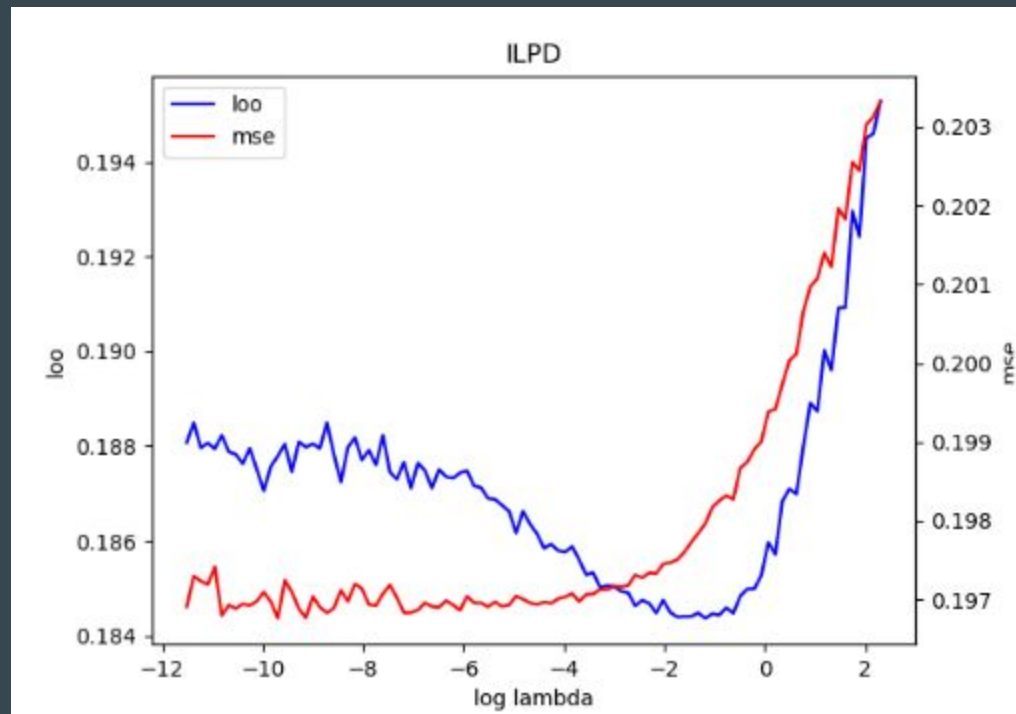
$$\sigma_i = \text{diag}(\mathbf{P})_i^{-1} (\mathbf{P} \mathbf{y})_i$$

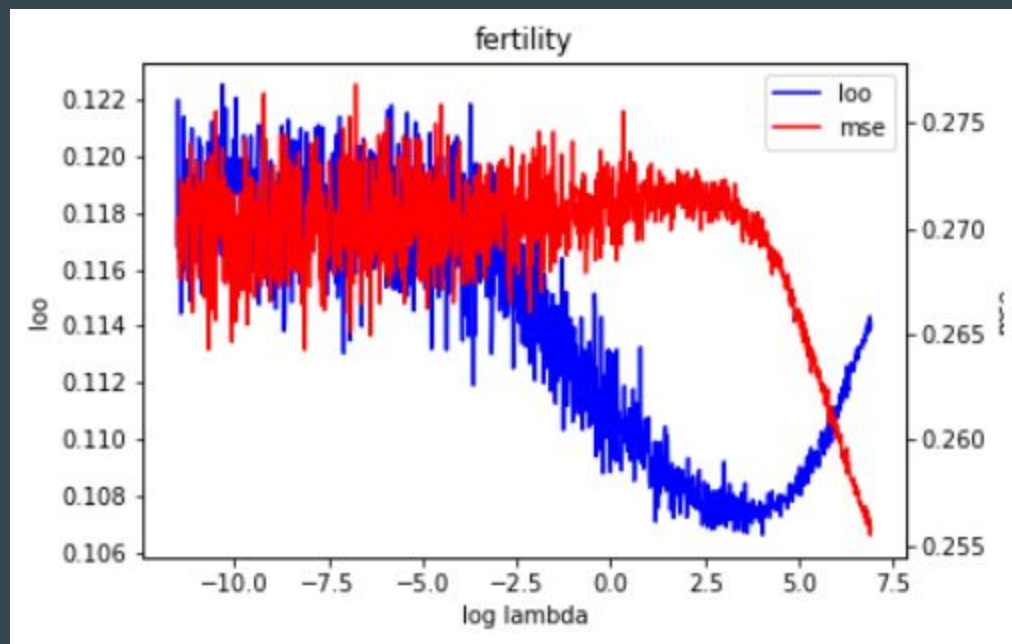
$$\text{LOO} = N^{-1} \boldsymbol{\sigma}^T \boldsymbol{\sigma}$$

$$C(\mathbf{H}) = \sum_i \sum_j h_i h_j$$





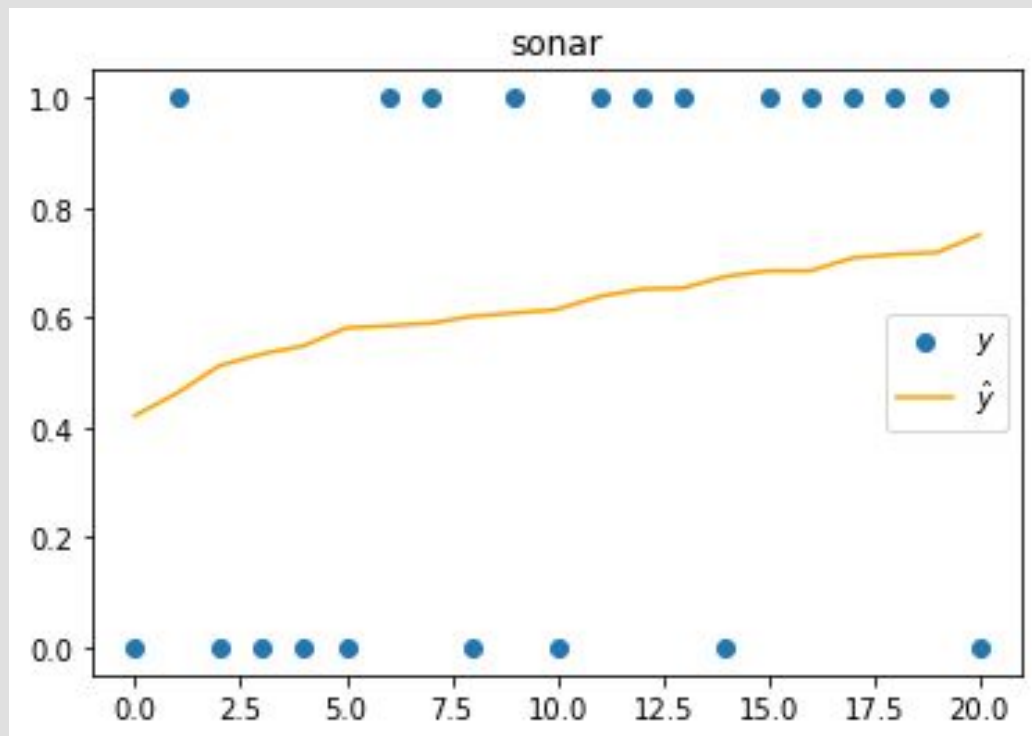


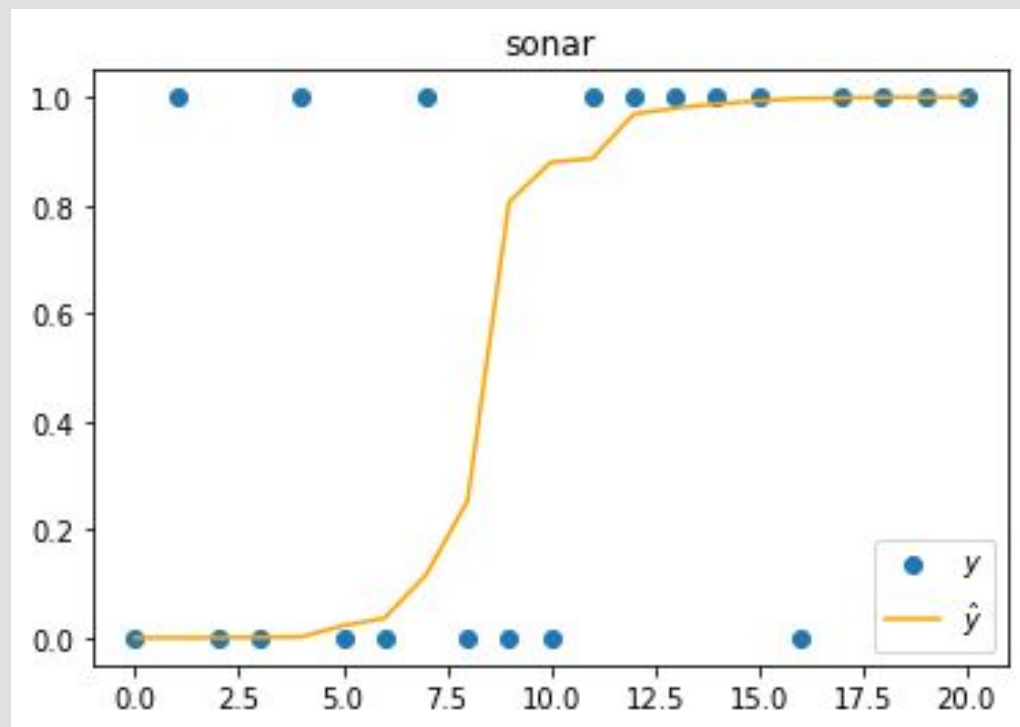


Resultados

Tabela III
RESULTADOS DE DESEMPENHO.

<i>Dataset</i>	MLP	AE	AE-C	RG
australian	0.93	0.92	0.93	0.86
banknote	1.00	1.00	1.00	0.97
breastcancer	0.99	1.00	1.00	0.97
breastHess	0.87	0.89	0.88	0.76
bupa	0.70	0.69	0.70	0.58
climate	0.96	0.91	0.96	0.77
diabetes	0.83	0.83	0.83	0.72
fertility	0.67	0.67	0.66	0.50
german	0.79	0.76	0.80	0.69
golub	0.71	0.69	0.60	0.77
haberman	0.69	0.70	0.72	0.50
heart	0.91	0.90	0.91	0.83
ILPD	0.75	0.72	0.72	0.52
parkinsons	0.89	0.87	0.88	0.76
sonar	0.88	0.84	0.81	0.79





Classes Desbalanceadas

Heurística de Amostragem

Grafo de Gabriel

$$Q(x_i) = \frac{V_{eq}(x_i)}{V(x_i)}$$

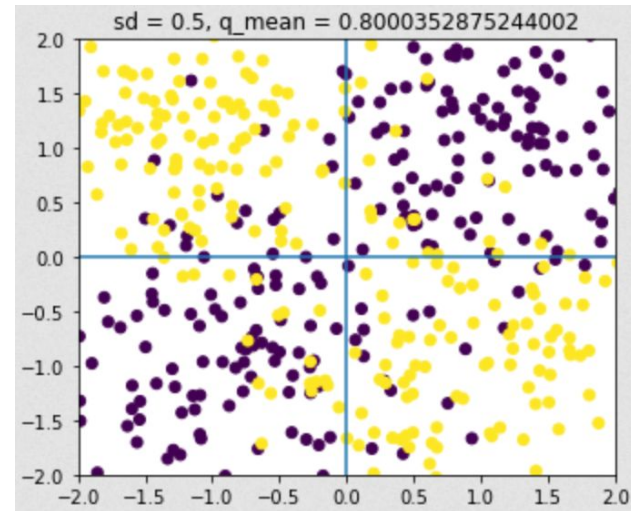


Tabela I
CARACTERÍSTICAS DAS BASES DE DADOS SELECIONADAS.

<i>Dataset</i>	Amostras	Atributos	η
australian	690	14	0.44
banknote	1372	4	0.44
breastcancer	683	6	0.65
breastHess	133	30	0.74
bupa	345	6	0.42
climate	540	18	0.91
diabetes	768	8	0.65
fertility	100	9	0.12
german	1000	24	0.70
golub	72	50	0.65
haberman	306	3	0.74
heart	270	13	0.56
ILPD	579	10	0.72
parkinsons	195	22	0.75
sonar	208	60	0.47

Tabela IV
CARACTERÍSTICAS DO CONJUNTO DE TREINAMENTO NA TÉCNICA DE
downsampling.

<i>Dataset</i>	<i>Amostras</i>	η
australian	427.8	0.61
banknote	586	0.94
breastcancer	243.1	0.12
breastHess	76.6	0.62
bupa	299.6	0.42
climate	476.6	0.91
diabetes	660.7	0.67
fertility	40.5	0.09
german	795.1	0.67
golub	52.6	0.58
haberman	169.2	0.71
heart	206	0.62
ILPD	362.9	0.65
parkinsons	83.9	0.49
sonar	178.8	0.44

Tabela V
RESULTADOS DE DESEMPENHO COM A HEURÍSTICA.

<i>Dataset</i>	MLP	AE	AE-C	RG
australian	0.93	0.92	0.93	0.86
banknote	1.00	1.00	1.00	0.97
breastcancer	0.99	1.00	1.00	0.97
breastHess	0.87	0.89	0.88	0.76
bupa	0.70	0.69	0.70	0.58
climate	0.96	0.91	0.95	0.77
diabetes	0.83	0.83	0.83	0.72
fertility	0.67	0.67	0.66	0.50
german	0.79	0.76	0.80	0.69
golub	0.71	0.69	0.63	0.77
haberman	0.69	0.70	0.71	0.50
heart	0.91	0.90	0.91	0.83
ILPD	0.75	0.72	0.72	0.52
parkinsons	0.89	0.87	0.88	0.76
sonar	0.88	0.84	0.81	0.79

Discussão

Discussão

- Representação não supervisionada se mostrou suficiente, destoante custo e saída não polarizada
- Exploração de novas funções de custo
 - Otimização multi-objetivo
- Modelo
 - VAE | GAN

Obrigado