Slicer: Feature Learning for Class Separability with Least-Squares Support Vector Machine Loss and COVID-19 Chest X-ray Case Study

David Charte Iván Sevillano María J. Lucena José L. Martín Francisco Charte Francisco Herrera

23 sep 2021 – HAIS 2021 (Deep Learning)

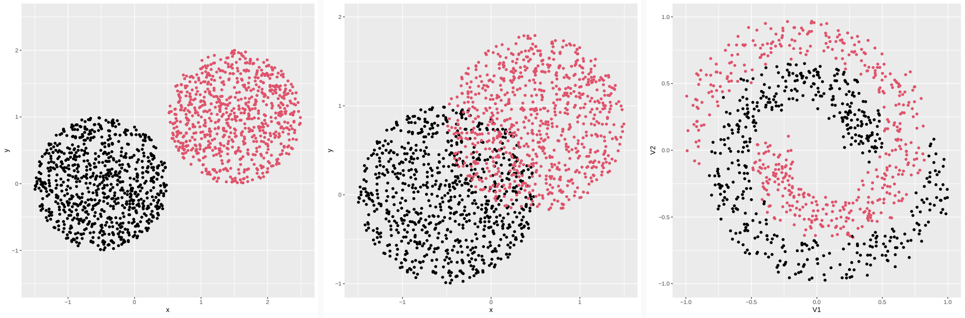
Andalusian Research Institute on Data Science and Computational Intelligence (UGR/UJA)
Hospital Universitario Clínico San Cecilio de Granada, Granada, Spain

UNIVERSIDAD DE GRANADA Universidad de Jaén HOSPITAL UNIVERSITARIO CLÍNICO SAN CECILIO DaSCI

Introduction

Introduction: data complexity

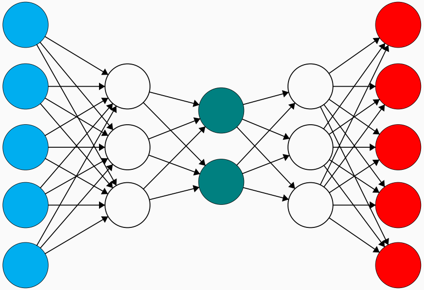
- Classes can have **overlap** or **complex boundaries**
- Other sources of complexity: dimensionality, imbalance, nonlinearity
- Complexity metrics: F1, F2, F3, N1, N2, …



Introduction: autoencoders

$$\text{Input space} \xrightarrow{\mathbf{f}} \text{Encoding space} \xrightarrow{\mathbf{g}} \text{Input space}$$

Optimize so that $\mathbf{g}(\mathbf{f}(\mathbf{x})) \approx \mathbf{x}$
(optionally, some other properties)



Objective

Learn simpler features to:

- improve the performance of basic/interpretable classifiers
- combine with variables from other sources

Slicer

Introducing class information for complexity reduction

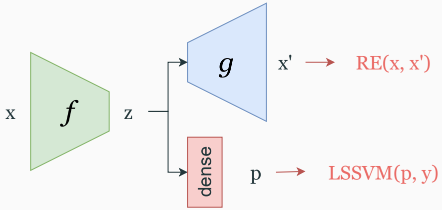


Figure 1: Basic schema of the Slicer model

$$\mathcal{J}_{\mathsf{LSSVM}}(x, y, \theta) = \frac{\mu}{2} w^T w + \frac{\zeta}{2} \sum_{i=1}^{n} \left(y^{(i)} - w^T f\left(x^{(i)}\right) + b \right)^2 \quad (1)$$

$$\mathcal{J}(x, y, \theta) = \mathcal{J}_{\mathsf{RE}}(x, \theta) + \mathcal{J}_{\mathsf{LSSVM}}(x, y, \theta) \quad (2)$$

Slicer model for feature learning in images

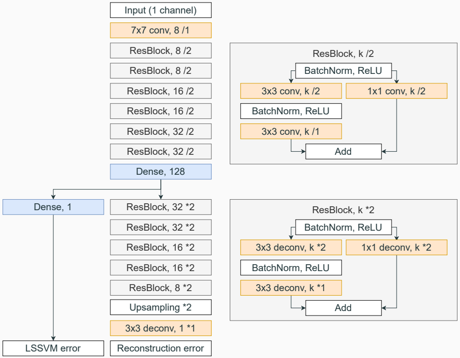


Figure 2: A convolutional autoencoder using Slicer loss

Experiments

COVIDGR-1 dataset

Openly available chest X-ray dataset for COVID detection:
https://github.com/ari-dasci/OD-COVIDGR

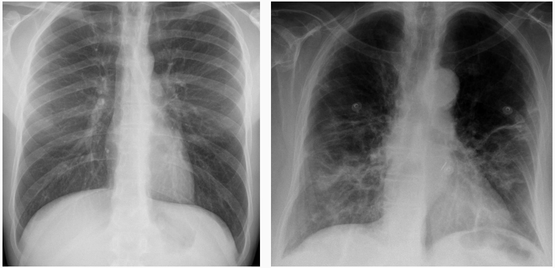


Figure 3: A negative sample and a positive one

class	#	severities	
negative	426		
positive	426	Normal-PCR+	76
		Mild	100
		Moderate	171
		Severe	79

Table 1: Class distribution. Normal-PCR+ = PCR-positive patients with 'asymptomatic' X-rays.

Experiment workflow

Dataset $\longrightarrow$ **Feature learner** $\longrightarrow$ Classifiers $\longrightarrow$ Evaluation metrics

Feature sets	Classifiers	Evaluation metrics	
Original	Decision tree (DT)	Accuracy	$\frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}$
Basic AE	k nearest neighbors (kNN)	Precision	$\frac{\text{TP}}{\text{TP}+\text{FP}}$
Slicer	Support vector machine (SVM)	Recall	$\frac{\text{TP}}{\text{TP}+\text{FN}}$
	Gaussian process (GP)	F1-score	$\frac{2\cdot\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}}$

Source code: https://github.com/fdavidcl/slicer-conv

Results

Classifier	Features	Accuracy	Precision	Recall	F1-score
DT	original	58.098	58.259	**57.470**	57.665
DT	autoencoder	58.377	**58.889**	55.919	57.196
DT	slicer	**58.593**	58.837	57.422	**57.988**
GP	original	50.024	45.333	1.266	2.452
GP	autoencoder	50.024	4.000	0.047	0.093
GP	slicer	**62.656**	**64.339**	**56.673**	**60.032**
kNN	original	**62.585**	**65.804**	52.780	58.358
kNN	autoencoder	61.837	65.091	51.469	57.194
kNN	slicer	62.326	63.624	**57.284**	**60.092**
SVM	original	**67.329**	66.611	**69.920**	**67.931**
SVM	autoencoder	67.072	**67.006**	67.622	67.057
SVM	slicer	65.987	66.235	65.025	65.393

Table 2: Average classification metrics over 25 runs (5x 5-fold CV).

GAN+CNN-based models achieve 76.18% accuracy / 75.71 F1 on this same dataset[a].

[a] Tabik, Siham, et al. COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on Chest X-Ray images. IEEE Journal of Biomedical and Health Informatics, 2020, vol. 24, no 12, p. 3595-3605.

Conclusion

Conclusions

- Slicer is an AE-based feature learner with linear classification loss for complexity reduction
- The model improves features over basic AE and maintains or improves respect to the original data

Future work:

- Analyze the influence of severity levels in class complexity
- Combine learned features with clinical data
- Feature disentanglement for easier interpretation
- Variational or adversarial versions of the model

Thanks for your attention!

Slicer: Feature Learning for Class Separability with Least-Squares Support Vector Machine Loss and COVID-19 Chest X-ray Case Study

David Charte Iván Sevillano María J. Lucena José L. Martín Francisco Charte Francisco Herrera

23 sep 2021 – HAIS 2021 (Deep Learning)

Andalusian Research Institute on Data Science and Computational Intelligence (UGR/UJA)
Hospital Universitario Clínico San Cecilio de Granada, Granada, Spain

UNIVERSIDAD DE GRANADA
Universidad de Jaén
HOSPITAL UNIVERSITARIO CLÍNICO SAN CECILIO
DaSCI