University of Liège
Faculty of Applied Sciences
Department of Electrical Engineering & Computer Science

PhD dissertation

Contributions to Deep Generative Modeling

From black-box to hybrid approaches

by Antoine Wehenkel

Advisor: Prof. Gilles Louppe
August 2022

JURY MEMBERS

???, Professor at the Université de Liège (President);

Gilles Louppe, Professor at the Université de Liège (Advisor);

???, Professor at the Université de Liège;

???, ???;

???, ???;

???, ???;

ABSTRACT

To be completed.

ACKNOWLEDGMENTS

To be completed.

CONTENTS

1 INTRODUCTION

If you erase this you are almost there! :D

1.1 OUTLINE AND CONTRIBUTIONS

If you erase this you are almost there! :D

1.2 PUBLICATIONS

To be completed.

2 BACKGROUND

Outline

This chapter concerns the definition of unsupervised learning with a brief review of classical methods. Graphical models (in particular B-net are introduced here.) It continues with a review of deep generative modeling, with a discussion between explicit and implicit generative modeling. We introduce the concepts of GANs, VAEs, Normalizing Flows and diffusion models. With a note that VAEs and diffusions models are discussed in more details in further chapters.

2.1 UNSUPERVISED LEARNING

Part I

BLACK-BOX GENERATIVE MODELING

3 DEEP LATENT VARIABLES GENERATIVE MODELS

Outline

???

3.1 VARIATIONAL AUTO ENCODERS

3.2 DIFFUSION MODELS

3.3 CONTRIBUTION: *diffusion priors in variational autoencoders*

4

UNSTRUCTURED DENSITY ESTIMATION WITH NORMALIZING FLOWS

Outline

???

4.1 NORMALIZING FLOWS

4.2 CONTRIBUTION: *you say normalizing flows i see bayesian networks*

4.3 CONTRIBUTION: *unconstrained monotonic neural networks*

5

DIGRESSION ON MONOTONIC FUNCTIONS IN MACHINE LEARNING

Outline

???

Part II

HYBRID GENERATIVE MODELING

6 STRUCTURED DENSITY ESTIMATION WITH NORMALIZING FLOWS

Outline
???

6.1 contribution: *graphical normalizing flows*

7 HYBRID GENERATIVE MODELS

Outline
???

7.1 CONTRIBUTION: *robust hybrid learning with expert augmentation*

Part III

APPENDIX

A

NOTATIONS

$\mathcal{A}$	A supervised learning algorithm .... **??**
$\mathcal{A}(\theta, \mathcal{L})$	The model $\varphi_{\mathcal{L}}$ produced by algorithm $\mathcal{A}$ over $\mathcal{L}$ and hyper-parameters θ .... **??**
α_s	The proportion of samples in a random patch .... **??**
α_f	The proportion of features in a random patch .... **??**
b_l	The l-th value of a categorical variable .... **??**
B	A subset $B \subseteq V$ of variables .... **??**
c_k	The k-th class .... **??**
C_p^k	The number of k-combinations from a set of p elements .... **??**
$C(N)$	The time complexity for splitting N samples ...... **??**
$\mathbb{E}$	Expectation ....
$\overline{E}(\varphi_{\mathcal{L}}, \mathcal{L}')$	The average prediction error of $\varphi_{\mathcal{L}}$ over $\mathcal{L}'$ .... **??**
$Err(\varphi_{\mathcal{L}})$	The generalization error of $\varphi_{\mathcal{L}}$ .... **??**, **??**
$H(X)$	The Shannon entropy of X .... **??**
$H(X\|Y)$	The Shannon entropy of X conditional to Y........ **??**
$\mathcal{H}$	The space of candidate models .... **??**
$i(t)$	The impurity of node t .... **??**, **??**
$i_R(t)$	The impurity of node t based on the local resubstitution estimate .... **??**, **??**
$i_H(t)$	The entropy impurity of node t .... **??**
$i_G(t)$	The Gini impurity of node t .... **??**
$\Delta i(s, t)$	The impurity decrease of the split s at node t ..... **??**
$I(X; Y)$	The mutual information between X and Y .... **??**
$Imp(X_j)$	The variable importance of X_j .... **??**, **??**
J	The number of classes .... **??**
K	The number of folds in cross-validation .... **??** The number of input variables drawn at each node for finding a split .... **??**
$K(\mathbf{x}_i, \mathbf{x}_j)$	The kernel of $\mathbf{x}_i$ and $\mathbf{x}_j$ .... **??**, **??**
L	A loss function .... **??** The number of values of a categorical variable .... **??**
$\mathcal{L}$	A learning set $(\mathbf{X}, \mathbf{y})$ .... **??**
$\mathcal{L}^m$	The m-th bootstrap replicate of $\mathcal{L}$ .... **??**

$\mathcal{L}_t$	The subset of node samples falling into node t	??
M	The number of base models in an ensemble	??
$\mu_{\mathcal{L},\theta_m}(\mathbf{x})$	The mean prediction at $X = \mathbf{x}$ of $\varphi_{\mathcal{L},\theta_m}$	??
N	The number of input samples	??
N_t	The number of node samples in node t	??
N_{ct}	The number of node samples of class c in node t	??
Ω	The universe, or population, from which cases are sampled	??
p	The number of input variables	??
p_L	The proportion of node samples going to t_L	??
p_R	The proportion of node samples going to t_R	??
$p(t)$	The estimated probability $p(X \in \mathcal{X}_t) = \frac{N_t}{N}$	??
$p(c\|t)$	The empirical probability estimate $p(Y = c\|X \in \mathcal{X}_t) = \frac{N_{ct}}{N_t}$ of class c at node t	??
$\widehat{p}_{\mathcal{L}}$	An empirical probability estimate computed from the learning set $\mathcal{L}$	??
$P(X, Y)$	The joint probability distribution of the input variables $X = (X_1, \dots, X_p)$ and the output variable Y	??
$\mathcal{P}_k(V)$	The set of subsets of V of size k	??
φ	A model or function $\mathcal{X} \mapsto \mathcal{Y}$	??
	A single decision tree	??
$\widetilde{\varphi}$	The set of terminal nodes in φ	??
$\varphi(\mathbf{x})$	The prediction of φ for the sample $\mathbf{x}$	??
$\varphi_{\mathcal{L}}$	A model built from $\mathcal{L}$	??
$\varphi_{\mathcal{L},\theta}$	A model built from $\mathcal{L}$ with random seed θ	??
φ_B	A Bayes model	??
$\psi_{\mathcal{L},\theta_1,\dots,\theta_M}$	An ensemble of M models built from $\mathcal{L}$ and random seeds $\theta_1, \dots, \theta_M$	??
$\mathcal{Q}$	A set $\mathcal{Q} \subseteq \mathcal{S}$ of splits of restricted structure	??, ??
$\mathcal{Q}(X_j)$	The set $\mathcal{Q}(X_j) \subseteq \mathcal{Q}$ of univariate binary splits that can be defined on variable X_j	??, ??
$\rho(\mathbf{x})$	The correlation coefficient between the predictions at $X = \mathbf{x}$ of two randomized models	??
s	A split	??, ??
s^*	The best split	??, ??
s_j^*	The best binary split defined on variable X_j	??, ??
s_j^v	The binary split $(\{\mathbf{x}\|x_j \leqslant v\}, \{\mathbf{x} > v\})$ defined on variable X_j with discretization threshold v	??
s_t	The split labeling node t	??
$\tilde{s}_t^j$	The best surrogate split for s_t defined from X_j	??

$\mathcal{S}$	The set of all possible splits s	??
$\sigma^2_{\mathcal{L},\theta_m}(\mathbf{x})$	The prediction variance at $X = \mathbf{x}$ of $\varphi_{\mathcal{L},\theta_m}$	??
t	A node in a decision tree	??
t_L	The left child of node t	??, ??
t_R	The right child of node t	??, ??
θ	A vector of hyper-parameter values	??
	A random seed	??
θ^*	The optimal hyper-parameters	??
$\widehat{\theta}^*$	The approximately optimal hyper-parameters	??
θ_m	The seed of the m-th model in an ensemble	??
v	A discretization threshold in a binary split	??
v_k	The k-th value of an ordered variable, when node samples are in sorted order	??
v'_k	The mid-cut point between v_k and v_{k+1}	??
V	The set $\{X_1, \dots, X_p\}$ of input variables	??
V^{-j}	$V \setminus \{X_j\}$	??
$\mathbb{V}$	Variance	
$\mathbf{x}$	A case, sample or input vector $(x_1, \dots, x_p)$	??
$\mathbf{x}_i$	The i-th input sample in $\mathcal{L}$	??
x_j	The value of variable X_j for the sample $\mathbf{x}$	??
$\mathbf{X}$	The $N \times p$ matrix representing the values of all N samples for all p input variables	??
X_j	The j-th input variable or feature	??, ??
X	The random vector $(X_1, \dots, X_p)$	??
$\mathcal{X}_j$	The domain or space of variable X_j	??
$\mathcal{X}$	The input space $\mathcal{X}_1 \times \dots \times \mathcal{X}_p$	??
$\mathcal{X}_t$	The subspace $\mathcal{X}_t \subseteq \mathcal{X}$ represented by node t	??
y	A value of the output variable Y	??
$\widehat{y}_t$	The value labelling node t	??
$\widehat{y}^*_t$	The optimal value labelling node t	??
$\mathbf{y}$	The output values $(y_1, \dots, y_N)$	??
Y	The output or response variable Y	??
$\mathcal{Y}$	The domain or space of variable Y	??

B

REFERENCES