Adversarial Training of Neural Networks against Systematic Uncertainty

Gilles Louppe
New York University
g.louppe@nyu.edu

Abstract

1 Introduction

[GL: Distinction between statistical and systematic uncertainty.] [GL: Define nuisance parameters.] [GL: We want to build an accurate classifier whose output remains invariant with respect to systematic uncertainties.]

2 Problem statement

Let assume a probability space $(\Omega, \mathcal{F}, P)$, where Ω is a sample space, $\mathcal{F}$ is a set of events and P is a probability measure. Let consider the multivariate random variables $X_\lambda : \Omega \mapsto \mathbb{R}^p$ and $Y : \Omega \mapsto \mathcal{Y}$, where X_λ depends on a nuisance parameter λ encoding for systematic uncertainties in its probability distribution. That is, X_λ and Y induce together the joint probability distribution $p(X_\lambda, Y)$. For training, let further assume a finite set $\{\lambda_i, x_i, y_i, \}_{i=1}^N$ of realizations $X_{\lambda_i}(\omega_i), Y(\omega_i)$, for $\omega_i \in \Omega$ and known values λ_i of the nuisance parameter. Our goal is to learn a function $f : \mathbb{R}^p \mapsto \mathcal{Y}$ (e.g., a classifier if $\mathcal{Y}$ is a finite set of classes) minimizing the expected value of a loss $L(Y, f(X_\lambda))$, with the constraint that $f(X_\lambda)$ should be robust to the value of the nuisance parameter λ – which remains unknown at test time. More specifically, we aim at building f such that in the ideal case

$$f(X_{\lambda_i}(\omega)) = f(X_{\lambda_j}(\omega)) \tag{1}$$

for any sample $\omega \in \Omega$ and any λ_i, λ_j pair of values of the nuisance parameter.

Since we do not have training tuples $(X_{\lambda_i}(\omega), X_{\lambda_j}(\omega))$ (for the same unknown ω), we propose instead to solve the closely related problem of finding a predictive function f such that

$$P(\{\omega | f(X_{\lambda_i}(\omega)) = y\}) = P(\{\omega' | f(X_{\lambda_j}(\omega')) = y\}) \text{ for all } y \in \mathcal{Y}. \tag{2}$$

In words, we are looking for a predictive function f such that the distribution of $f(X_\lambda)$ is invariant with respect to the nuisance parameter λ. Note that a function f for which Eqn. 1 is true necessarily satisfies Eqn. 2. The converse is however in general not true, since the sets of samples $\{\omega | f(X_{\lambda_i}(\omega)) = y\}$ and $\{\omega' | f(X_{\lambda_j}(\omega)) = y\}$ do not need to be the same for the equality to hold. [GL: Yet, in practice, this criterion is often good enough.]

3 Method

4 Experiments

5 Related work

[GL: Similar to domain adaptation, but with infinitely many domains, as parameterized by λ.]

29th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

6 Conclusions

Acknowledgments