

[preprint,12pt,authoryear]elsarticle [utf8]inputenc hyperref url booktabs amsfonts amsmath amsthm geometry nicefrac microtype xcolor svg xcolor,listings [ide]sty/listings-ceylon syntax sty/bcprules tikz multicol arrows.meta appendix pgfplots arydshln subcaption algorithm algpseudocode draftwatermark 0.95 25 5 5cm DRAFT

pdfpages mathrsfs arrows [patterns] shapes.geometric decorations.pathreplacing,matrix table/.style=matrix of math nodes, row sep=-, column sep=-, nodes=rectangle,text width=1em,align=center, text depth=1.ex, text height=1ex, nodes in empty cells, left delimiter=[, right delimiter=], ampersand replacement=& minted caption subcaption

remarkRemark defiDefinition VocabularyVocabulary notationNotation defoDefinition exExample theoremTheorem[section] arg $\max \arg \min \nabla \theta_{J_k} \nabla \theta_m [1] 1 [1] 1$ |*gradient estimator for symbolical features* GSEone – hot – end

[2]12 $\sum [1] E[1][1]$ *equation 1*

xcolor, soul lightblue rgb0.9, 0.9, 1.0 fffqffrgb1,0,1 lightblue

document frontmatter

Stochastic Gradient Descent with

[Lokad]organization=Lokad, city=Paris, country=FRANCE [Litis]organization=Litis, city=Rouen, country=FRANCE [Lokad,Litis]Paul PeseuxCorauthor [Litis]Maxime Berar [Litis]Thierry Paquet [Lokad]Victor Nicollet

[Corauthor]corresponding author, paul.peseux@lokad.com

abstract

Symbolic data are present in key areas such as health or supply chain, and this data require specific treatment. In order to apply recent machine learning models on such data, encoding is needed. In order to build interpretable models, is a very good solution. But such encoding creates sparse data. Gradient estimators are not suited for sparse data: the gradient is mainly considered as zero while it simply does not always exists. After a survey on symbolic data in public datasets, a novel gradient estimator is introduced. We show what this estimator minimizes in theory and show its efficiency on 6 different datasets with multiple model architectures.

This new estimator performs better than common estimators under similar settings. Overall, the aim of this paper is to convince researchers to thoroughly consider symbolic data and adapt their models, optimizers, benchmarks ... to these key features.