Information in Local Curvature: Three Papers on Adaptive Methods in Computational Statistics

by

Berent Ånund Strømnes Lunde

Thesis submitted in fulfilment of
the requirements for the degree of
PHILOSOPHIAE DOCTOR
(PhD)

Faculty of Science and Technology
Department of Mathematics and Physics
2020

University of Stavanger
NO-4036 Stavanger
NORWAY
www.uis.no

ISBN:
ISSN:
PhD: Thesis UiS No.

Preface

This thesis is submitted in partial fulfilment of the requirements for the degree of Philosophiae Doctor (PhD) at the University of Stavanger, Faculty of Science and Technology, Norway. The research has been carried out at the University of Stavanger from September 2017 to August 2020.

The present work is divided in two. The first part gives a brief introduction and background information to the most important topics and ideas of the work. The second part consists of the following papers:

Paper I

Lunde, Berent Ånund Strømnes, Tore Selland Kleppe, and Hans Julius Skaug (2020). Saddlepoint adjusted inversion of characteristic functions. *Journal of Journals* 57, 80-93.

Paper II

Lunde, Berent Ånund Strømnes, Tore Selland Kleppe, and Hans Julius Skaug (2020). An information criterion for automatic gradient tree boosting. *To be submitted for publication in Journal of Journals.*

Paper III

Lunde, Berent Ånund Strømnes, Tore Selland Kleppe, and Hans Julius Skaug (2020). aGTBoost: Adaptive and Automatic Gradient Tree Boosting Computations. *To be submitted for publication in Journal of Journals.*

Acknowledgements

I would like to thank my supervisor, Professor Tore Selland Kleppe, for his constant support and invaluable guidance. You have allowed me to pursue ideas, to fail, and so many times steered me in the right direction with detailed feedback and questions that would uncover flaws, but which would eventually lead me closer towards truth. Thank you for your encouraging words, enthusiasm and genuine thoughtfulness.

Thanks are also due to my co-supervisors, the professors Hans Julius Skaug and Jan Terje Kvaløy. Professor Skaug has co-authored two of the papers in this thesis, allowed a stay at the University of Bergen, and always shared of his time and wide experience, for this I am grateful. Professor Kvaløy has been presence of constant cheerfulness and inspiration. Thank you for always taking a genuine interest in people and their ideas. I extend my thanks to my fellow PhD-students. Utmost appreciation goes out to Kjartan Kloster Osmundsen and Birthe Aarekol, for the many discussions and several trips all around the world to academic conferences and meetings.

Finally, I want to give many thanks to my friends, in particular Kjetil for his unlimited accommodating spirit. And also my family, in particular my mother, Katrin, for invaluable advice, and my wife, Saeron Min, which has been a force of continued support. You have let me dive into hours of silent thoughts, calculations and coding when it was needed, but have also pulled me away and forced upon me a more balanced life when I would encounter a runtime error of the mind, but of the kind which I would not be able to see or solve by myself.

Berent Ånund Strømnes Lunde
Stavanger, August 2020

Abstract

Advanced statistical computations have become increasingly important, as with the increased flexibility of models capturing complex relationships in new data and use-cases, comes increased difficulties of fitting procedures for the models. For example if the model is complex, involving multiple sources of randomness, then the probability density function used in maximum likelihood estimation typically does not have a closed form. On the other hand, in regression type problems the closed form of the conditional distribution of the response is often known. However, the relationship between features and response can be complex, high dimensional and is generally unknown, motivating non-parametric procedures with new sets of fitting problems.

This thesis explores techniques utilizing the local curvature of objective functions, and using the information inherent in this local curvature, to create more stable and automatic fitting procedures. In the first paper, a saddlepoint adjusted inverse Fourier transform is proposed. The method performs arbitrarily accurate numerical inversion, even in the tails of the distribution. This allow practitioners to specify their models in terms of the characteristic function which often exists in closed form. The second paper proposes an information criterion for the node-splits, after greedy binary splitting, in gradient boosted trees. This removes the need for computationally expensive cross validation and expert opinions in tuning hyperparameters associate gradient tree boosting. The third paper focuses on the implementation of the theory presented in the second paper into the R-package aGTBoost, and also builds on the information criterion to suggest an adjustment of ordinary greedy-binary-splitting, adapted to gradient tree boosting.

Table of Contents

Preface iii

Acknowledgements iv

Abstract vi

1 Introduction 1

2 Maximum likelihood and supervised learning 3
- 2.1 Maximum likelihood estimation 3
- 2.2 Supervised learning 3

3 Quadratic approximations in statistics 5
- 3.1 The saddlepoint approximation 6
- 3.2 Gradient tree boosting 8
- 3.3 First order asymptotics 10
- 3.4 Model selection 10

4 Summary of the papers 11

References 13

Appendix

Saddlepoint adjusted inversion of characteristic functions 15

An information criterion for automatic gradient tree boosting 18

aGTBoost: Adaptive and Automatic Gradient Tree Boosting Computations .. 20

1 Introduction

Advanced statistical methods and procedures have seen increased and widespread usage in later years. This is backed by access to more data and new use-cases, cheaper computational power, and adaption into mainstream languages such as Python and R. Underlying this trend is also the increased usability of said algorithms, in regards to training on data and putting them into production. The goal of this thesis is to further the usability of computational methods in statistics with regards to stability, speed, and automatic functionality.

The main approach of the present work is to, in some loose and wide sense, approximate some objective function with a local quadratic approximation to either solve stability issues, create dynamic step-lengths, or measure the uncertainty of estimators. The hope is then that the iterative methods that the local quadratic approximation is applied to, will see an increased adaptivity to individual data and problems, and corresponding decrease in manual tuning performed before applications to the problem at hand.

The first part of this thesis will give a brief and informal introduction to the concepts and techniques that are used in papers I-III. The basis is the objective of maximum likelihood and supervised learning, which are presented in the first section. The second section introduces the local quadratic approximation, and showcase it in the relevant use-cases for papers I-III, i.e. maximum-likelihood numerical optimization, the saddlepoint approximation, gradient tree boosting, asymptotic theory and model selection. The final section summarises the papers of the thesis.

2 Maximum likelihood and supervised learning

Maximum likelihood estimation and supervised learning are briefly introduced in an informal manner. This is done to provide intuition to the fundamental objectives of the algorithms that are presented, and as motivation to the research on the algorithms's problems presented in this thesis.

2.1 Maximum likelihood estimation

Let $\mathbf{x}$ denote an n-dimensional vector of observations from a parametric distribution, with density denoted $p(\mathbf{x}; \theta_0)$, where $\theta_0 \in \Theta$, $\Theta \subseteq \mathrm{R}^p$ is a p-dimensional vector. Is often the case that a reasonable parametric family of functions, $p(\mathbf{x}; \theta)$, $\theta \in \Theta$, can be inferred from the problem and from inspection of the data. However, θ_0 will be unknown, and it is reasonable to estimate it using the observed data $\mathbf{x}$. To this end, maximum likelihood estimation is a popular approach. The maximum likelihood estimate (MLE) is the value of θ in Θ which maximizes the probability of the data, i.e. the likelihood,

$$\hat{\theta} = \arg\min_{\theta} \{-\log p(\mathbf{x}; \theta)\}. \tag{2.1}$$

The maximum likelihood estimate, $\hat{\theta}$ is, under suitable regularity conditions, the asymptotically unbiased minimum variance estimate, and asymptotically normal. See -Van der vaart- for a treatment of their asymptotic properties.

2.2 Supervised learning

The supervised learning objective is perhaps easiest stated as "regression", but also bears resemblance to maximum likelihood estimation. Assume now that $\mathbf{x} \in R^{n \times m}$ is a matrix of p covariates or features for n observations. Let $y \in R^n$ be an n-vector of response observations.

In general, individual response observations, $y_i,\ i = 1 \ldots n$, could also be multidimensional, but throughout this thesis they are assumed one-dimensional. Let $\hat{y}_i$ be a prediction for y_i and let the loss function $l(y_i, \hat{y}_i)$ be a function measuring the difference between a response and its prediction. The supervised learning objective is to find the best possible predictive function, $f(x) = \hat{y}$, which takes a feature vector (row-vector of $\mathbf{x}$) as its argument, and outputs a prediction $\hat{y}$. "Best possible" is here in reference to the loss l over observations not part of the training data $(\mathbf{x}, y)$. More formally, we seek f so that

$$\hat{f} = \arg\min_{f} \left\{ E\left[l(y^0, f(x^0)) \right] \right\}, \tag{2.2}$$

where the superscript (y^0, x^0) indicates an observation unseen in the training data, and E denotes the expectation. Notice that, if the search is constrained over a parametric family of functions indexed by $\theta \in \Theta$, and the loss function is taken to be the negative log-likelihood, $l = -\log p$, then the supervised learning objective is closely related to the objective of maximum likelihood estimation (2.1) in a regression setting. In fact, the objective in (2.1) is the sample estimator of the expected value in (2.2), but biased downwards in expectation, as evaluation is done over observations in the training set.

3 Quadratic approximations in statistics

The maximum likelihood objective (2.1) and supervised learning objective (2.2) are, except for the most trivial of cases, not straightforward, and must be solved numerically. This then typically involve some iterative algorithm, which may require substantial manual tuning and trial and error before successful application. However, a local quadratic approximations to some otherwise intractable function can often be of help in making these algorithms more automatic and adaptive to the data and problem at hand.

When referring to a local quadratic approximation, as is frequently done in this thesis, it is meant to refer to a 2'nd order Taylor approximation of a function $f(x)$, about some point x_0. For example, the quadratic approximation of the negative log-likelihood loss function $l = -\log p$ about some value of θ, say θ_k, gives

$$\begin{aligned} l(y_i, f(x_i;\theta)) \approx\ & l(y_i, f(x_i;\theta_k)) + \nabla_\theta l(y_i, f(x_i;\theta_k))(\theta - \theta_k) \\ & + \frac{1}{2}(\theta - \theta_k)^T \nabla^2_\theta l(y_i, f(x_i;\theta_k))(\theta - \theta_k). \end{aligned} \tag{3.1}$$

Example 3.0.1 (Newton-Raphson) *The MLE $\hat{\theta}$ in* (2.1) *typically has to be found numerically, as the score equations, $0 = \nabla_\theta l(y_i, f(x_i;\theta_k))$, are not possible to solve analytically. Assuming that l is differentiable and convex in θ, the Newton-Raphson algorithm will converge to the MLE $\hat{\theta}$. The iterative Newton-Raphson algorithm is constructed by employing the r.h.s. of* (3.1) *iteratively to the current value of θ, say θ_k, the next value in the iterative algorithm is then given by*

$$\theta_{k+1} = \theta_k - \left[\nabla^2_\theta l(y_i, f(x_i;\theta_k))\right]^{-1} \nabla_\theta l(y_i, f(x_i;\theta_k)),$$

the MLE if l indeed was equal to the quadratic approximation on the r.h.s. in (3.1).

There are many problems in computational statistics that may be helped by (3.1), however only a few, the ones relevant to papers I-III, are

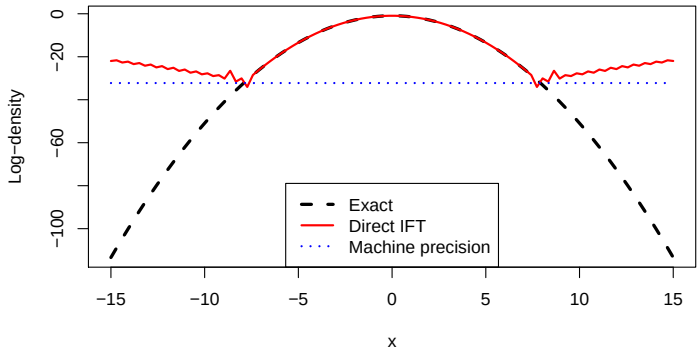

Figure 3.1: Figure included from Paper 1. Illustrating the dominance of inaccuracies of the IFT (3.2), calculated with quadrature, at machine precision at $\log(1.0 \times 10^{-14})$ indicated by the dotted horizontal line.

discussed here. The following sections discuss applications of local quadratic approximations as helpful tools in dealing with some of the problem associate/in dealing numerical optimization of (2.1) and (2.2).

3.1 The saddlepoint approximation

It is often the case that the density $p_X(x;\theta)$ of a random variable X, is not available in closed form when there are multiple sources of randomness present in X. Direct optimization of (2.1) is therefore difficult. However, the characteristic function or the Fourier transform of the density, $\varphi_X(s) = E[\exp(isX)]$ often still exhibits closed form, even in situations with more than one source of randomness. The density might then be retrieved by numerically evaluating the inverse Fourier transform

$$p_X(x;\theta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi_X(s;\theta)e^{-isx}ds = \frac{1}{2\pi}\int_{\infty}^{\infty} e^{K_X(is;\theta)-isx}ds, \quad (3.2)$$

where $K_X(s;\theta) = \log \varphi_X(-is;\theta)$ is the cumulative generating function (CGF).

However, consider the case of numerical MLE optimization, for example by using the Newton-Raphson algorithm in Example 3.0.1. Here, at the first iteration, say θ_1, the initial estimate is likely to start at values far from the population MLE, θ_0. Necessarily, observations x will take place in low-density regions of $p(x;\theta_1)$, and this will continue to be the case at subsequent iterations, until θ_k is close to θ_0. This constitutes a problem to direct numerical inversion of (3.2) using quadrature schemes (weighted sum of integrand evaluations), as numerical inaccuracies related to the (binary) representation of floating-point numbers will dominate. More specifically, considering double precision at order 1.0×10^{-16}, if x is in a region with log-density $\log p(x;\theta_k)$ smaller than this value, the inaccuracy of the representation is sure to dominate. Even more is that such behaviour/pathologies in practice happens a few orders of magnitude higher than the theoretical limit given above. In Figure 3.1, the error dominates already at 1.0×10^{-14}.

An inversion technique that does not suffer from erroneous computations in low-density regions, and in fact is renown for its tail-accuracy, is the saddlepoint approximation (SPA) -cite daniels, Butler-. It is developed in paper 1 through an argument of exponential tilting, which takes place on the "time-domain" side of the Fourier transform. Complimentary, an argument on the "frequency-domain" side is given here, that closely follows the derivation in -citet Butler, chapter-. First, notice that the value of the integral is unchanged if we integrate through a line parallel to the imaginary axis, say τ,

$$p_X(x;\theta) = \frac{1}{2\pi} \int_{\infty}^{\infty} e^{K_X(\tau+is;\theta)-(\tau+is)x} ds \tag{3.3}$$

Now, apply the quadratic approximation (3.1) to the log-integrand, $K_X(\tau + is;\theta) - (\tau + is)x$, locally about the value of τ solving the saddlepoint equations

$$\hat{\tau} = \arg\min_{\tau}\{K_X(\tau;\theta) - \tau x\}, \tag{3.4}$$

henceforth called the saddlepoint. This then gives the approximation of the log-integrand

$$K_X(\hat{\tau} + is; \theta) - (\hat{\tau} + is)x \approx K_X(\hat{\tau}) - \hat{\tau}x - \frac{1}{2}\frac{d^2}{d\tau^2}K_X(\hat{\tau})s^2. \tag{3.5}$$

Inserting this into the integral, and performing the transformation $u = \sqrt{\frac{d^2}{d\tau^2}K_X(\hat{\tau})}s$, then gives the SPA as

$$\begin{aligned} p_X(x;\theta) &\approx \frac{\exp(K_X(\hat{\tau}) - \hat{\tau}x)}{2\pi\sqrt{\frac{d^2}{d\tau^2}K_X(\hat{\tau})}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2} du \\ &= \frac{\exp(K_X(\hat{\tau}) - \hat{\tau}x)}{\sqrt{2\pi\frac{d^2}{d\tau^2}K_X(\hat{\tau})}} = spa_X(x;\theta). \end{aligned} \tag{3.6}$$

The SPA (3.6) is often accurate, and is asymptotically exact in n if there is some asymptotic normality underlying X, for example for $X = n^{-1}\sum_i X_i$. In particular, its relatively fast computation if implemented using automatic-differentiation software to solve the inner problem (3.4), and its previously mentioned tail-accuracy, are properties that are highly tractable. On the down-side, the SPA does not integrate to one, except for in a few special cases. Also, the approximation is often unimodal even if the target density is multimodal, which could very well be the case when X consists of multiple sources of randomness. A common technique is to multiply the SPA with a constant value c, where $c^{-1} = \int spa_X(x;\theta)dx$, that ensure it is a density. This is immediately more computationally costly, require bespoke implementation, and does not solve the problems of unimodality. These problems are the subject of Paper I.

3.2 Gradient tree boosting

The idea behind gradient boosting emerges as a numerical approach to the optimization problem in (2.2), similarly to the Newton-Raphson algorithm from Example 3.0.1 solving the optimization problem in

(2.1): Given an initial function $f^0(x) = f_0(x)$, one ideally seeks a function $f_1(x)$ minimizing

$$\hat{f}_1(x) = \arg\min_{f_1} E\left[l\left(y, f^0(x) + f_1(x)\right)\right]. \tag{3.7}$$

A reasonable substitute to $\hat{f}_1$ is to find the functional derivative of this objective and add the negative direction to the model, say $f^1 = f_0 + f_1$, and then repeat the procedure until convergence at iteration K, which would yield the final model $\hat{f} = f^0 + \cdots + f^K$.

Difficulties arise to this rather ideal procedure, as the joint distribution of (y, x) is generally unknown. The immediate solution, if having access to a dataset $\mathcal{D}_n = \{y_i, x_i\}_{i=1}^n$ of independent observations, is to average the loss over these observations, and instead, at iteration k, seek

$$\hat{f}_k(x) = \arg\min_{f_k} \frac{1}{n} \sum_{i=1}^{n} l\left(y_i, f^{k-1}(x_i) + f_k(x_i)\right). \tag{3.8}$$

Necessarily, the expectation cannot be computed explicitly and neither can the functional derivative.

Gradient boosting emerges as the collection of the above-mentioned ideas: Compute derivative information or pseudo-residuals, and fit a statistical model to these observations. Next, shrink the model to make space for new models.

The search among all possible functions have been constrained to some statistical model.

might hope to find the functional derivative of the objective in (2.2) with respect to f, say $-f^1$, and add the negative direction to the initial function, obtaining $f^0 + f^1$,

and then repeat the procedure until convergence at iteration K which yields the final model $\hat{f}^K = f_0 + \cdots + f_K$. However, as the joint distribution of (y, x) is generally unknown, the expectation cannot be computed

explicitly and neither can the functional derivative. Gradient boosting solves this by iteratively computing derivatives from the objective, and fitting a statistical model to this derivative information.

Viewing gradient boosting as functional gradient descent is correct in the sense that each f_k approximates some functional gradient.

First, the expectation is approximated by averaging, and the objective is approximated by a Taylor expansion

This then minimizes an approximation to the original objective.

3.3 First order asymptotics

3.4 Model selection

4 Summary of the papers

The first paper of the thesis, Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce iaculis nulla in est pellentesque ultrices. Nulla ex eros, iaculis vel malesuada vitae, iaculis eu mauris. Nullam sed bibendum erat. Cras tristique augue quis risus egestas porta. Aenean gravida quam congue, malesuada massa nec, vestibulum neque. Suspendisse potenti. Pellentesque non enim efficitur, lobortis ex ac, pulvinar sem. Cras sed velit tincidunt, pharetra elit a, pretium nunc. Donec imperdiet sem a nunc porttitor, non elementum ipsum blandit. Vivamus imperdiet ipsum quis justo gravida blandit. Vivamus sed mauris sed urna maximus pretium. Donec luctus placerat turpis. Praesent dui augue, viverra ultrices lacinia ac, efficitur quis nisi. Nullam pulvinar orci eu lacus mollis, molestie placerat tortor posuere.

In the second paper, Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce iaculis nulla in est pellentesque ultrices. Nulla ex eros, iaculis vel malesuada vitae, iaculis eu mauris. Nullam sed bibendum erat. Cras tristique augue quis risus egestas porta. Aenean gravida quam congue, malesuada massa nec, vestibulum neque. Suspendisse potenti. Pellentesque non enim efficitur, lobortis ex ac, pulvinar sem. Cras sed velit tincidunt, pharetra elit a, pretium nunc. Donec imperdiet sem a nunc porttitor, non elementum ipsum blandit. Vivamus imperdiet ipsum quis justo gravida blandit. Vivamus sed mauris sed urna maximus pretium. Donec luctus placerat turpis. Praesent dui augue, viverra ultrices lacinia ac, efficitur quis nisi. Nullam pulvinar orci eu lacus mollis, molestie placerat tortor posuere.

The third paper, Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce iaculis nulla in est pellentesque ultrices. Nulla ex eros, iaculis vel malesuada vitae, iaculis eu mauris. Nullam sed bibendum erat. Cras tristique augue quis risus egestas porta. Aenean gravida quam congue, malesuada massa nec, vestibulum neque. Suspendisse potenti. Pellentesque non enim efficitur, lobortis ex ac, pulvinar sem. Cras sed velit tincidunt, pharetra elit a, pretium nunc. Donec imperdiet sem a

nunc porttitor, non elementum ipsum blandit. Vivamus imperdiet ipsum quis justo gravida blandit. Vivamus sed mauris sed urna maximus pretium. Donec luctus placerat turpis. Praesent dui augue, viverra ultrices lacinia ac, efficitur quis nisi. Nullam pulvinar orci eu lacus mollis, molestie placerat tortor posuere.

Appendix

Saddlepoint adjusted inversion of characteristic functions 15

An information criterion for automatic gradient tree boosting 18

aGTBoost: Adaptive and Automatic Gradient Tree Boosting Computations..20

Paper I

Saddlepoint adjusted inversion of characteristic functions

The First Paper

Ole Olesen[1], Geir Geirsen[1], and Jens Jensen[2]

[1]Department of Mathematics and Physics, University of Stavanger, Norway
[2]Department of Safety, Economics and Planning, University of Stavanger, Norway

January 6, 2020

Abstract

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum

1 Introduction

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam eget interdum mauris. Suspendisse ut odio sed urna molestie sollicitudin at sit amet nisl. Vivamus vel libero vel metus eleifend pulvinar ac eget eros. Maecenas lobortis sem nisl, sed ultrices metus elementum eget. Cras leo urna, rhoncus a volutpat at, dapibus in eros. Sed sit amet nulla placerat, vulputate diam non, tempus nisl. Sed quis nibh ut lectus accumsan mollis ut et mi. Ut et egestas mauris. Etiam vel pharetra enim. Donec a laoreet justo, eu ultrices enim.

Nunc maximus eu quam in iaculis. Nulla eget metus et sem varius aliquam. Morbi a ipsum aliquet, vestibulum quam eget, commodo tortor. Praesent purus orci, fermentum at tellus eu, aliquam consectetur mauris. Nullam quis diam ligula. Etiam at ipsum vel diam hendrerit ornare quis ut lectus. Donec elementum at ipsum vel ullamcorper. Etiam in magna non mauris lobortis semper ac a leo. Nam ultricies purus a pretium ultrices. Donec a mi vitae leo egestas imperdiet. Vestibulum faucibus urna suscipit tincidunt suscipit. Duis id velit nec magna eleifend facilisis pellentesque dapibus nunc. Integer interdum nunc orci, a dictum magna pulvinar vitae. Nam ac maximus felis.

Paper II

An information criterion for automatic gradient tree boosting

The Second Paper

Ole Olesen[1], Geir Geirsen[1], and Jens Jensen[2]

[1]Department of Mathematics and Physics, University of Stavanger, Norway
[2]Department of Safety, Economics and Planning, University of Stavanger, Norway

January 6, 2020

Abstract

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum

1 Introduction

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam eget interdum mauris. Suspendisse ut odio sed urna molestie sollicitudin at sit amet nisl. Vivamus vel libero vel metus eleifend pulvinar ac eget eros. Maecenas lobortis sem nisl, sed ultrices metus elementum eget. Cras leo urna, rhoncus a volutpat at, dapibus in eros. Sed sit amet nulla placerat, vulputate diam non, tempus nisl. Sed quis nibh ut lectus accumsan mollis ut et mi. Ut et egestas mauris. Etiam vel pharetra enim. Donec a laoreet justo, eu ultrices enim.

Nunc maximus eu quam in iaculis. Nulla eget metus et sem varius aliquam. Morbi a ipsum aliquet, vestibulum quam eget, commodo tortor. Praesent purus orci, fermentum at tellus eu, aliquam consectetur mauris. Nullam quis diam ligula. Etiam at ipsum vel diam hendrerit ornare quis ut lectus. Donec elementum at ipsum vel ullamcorper. Etiam in magna non mauris lobortis semper ac a leo. Nam ultricies purus a pretium ultrices. Donec a mi vitae leo egestas imperdiet. Vestibulum faucibus urna suscipit tincidunt suscipit. Duis id velit nec magna eleifend facilisis pellentesque dapibus nunc. Integer interdum nunc orci, a dictum magna pulvinar vitae. Nam ac maximus felis.

Paper III

aGTBoost: Adaptive and Automatic Gradient Tree Boosting Computations

The Third Paper

Ole Olesen[1], Geir Geirsen[1], and Jens Jensen[2]

[1]Department of Mathematics and Physics, University of Stavanger, Norway
[2]Department of Safety, Economics and Planning, University of Stavanger, Norway

January 6, 2020

Abstract

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum

1 Introduction

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam eget interdum mauris. Suspendisse ut odio sed urna molestie sollicitudin at sit amet nisl. Vivamus vel libero vel metus eleifend pulvinar ac eget eros. Maecenas lobortis sem nisl, sed ultrices metus elementum eget. Cras leo urna, rhoncus a volutpat at, dapibus in eros. Sed sit amet nulla placerat, vulputate diam non, tempus nisl. Sed quis nibh ut lectus accumsan mollis ut et mi. Ut et egestas mauris. Etiam vel pharetra enim. Donec a laoreet justo, eu ultrices enim.

Nunc maximus eu quam in iaculis. Nulla eget metus et sem varius aliquam. Morbi a ipsum aliquet, vestibulum quam eget, commodo tortor. Praesent purus orci, fermentum at tellus eu, aliquam consectetur mauris. Nullam quis diam ligula. Etiam at ipsum vel diam hendrerit ornare quis ut lectus. Donec elementum at ipsum vel ullamcorper. Etiam in magna non mauris lobortis semper ac a leo. Nam ultricies purus a pretium ultrices. Donec a mi vitae leo egestas imperdiet. Vestibulum faucibus urna suscipit tincidunt suscipit. Duis id velit nec magna eleifend facilisis pellentesque dapibus nunc. Integer interdum nunc orci, a dictum magna pulvinar vitae. Nam ac maximus felis.