AFRREV STECH
An International Journal of Science and Technology
Bahir Dar, Ethiopia

Vol.1 (3) August-December, 2012:81-91

ISSN 2225-8612 (Print) ISSN 2227-5444 (Online)

Higher-Order Hybrid Gaussian Kernel in Meshsize Boosting Algorithm

Ishiekwene, C. C.
Department of Mathematics, University of Benin, Nigeria
E-mail: cycigar@yahoo.co.uk

Afere, B.A.E
Department of Mathematics & Statistics
Federal Polytechnic Idah, Nigeria

Abstract

In this paper, we shall use higher-order hybrid Gaussian kernel in a meshsize boosting algorithm in kernel density estimation. Bias reduction is guaranteed in this scheme like other existing schemes but uses the higher-order hybrid Gaussian kernel instead of the regular fixed kernels. A numerical verification of this scheme is conducted and the results are compared with the regular fixed kernels.

Key words: Boosting, kernel density estimates, bias reduction, higher-order hybrid Gaussian kernel, meshsize, noises

Introduction

Schapire in 1990 proposed the first boosting algorithm. Other authors like Freund (1995), Schapire and Singer (1999) but to mention a few have also made contributions. Boosting is a means of improving the performance of a ‘weak learner’. It is applied in this context using the higher-order hybrid Gaussian kernel. Boosting does not only guarantee an error rate that is better than random guessing but also deals with the correction of ‘noises’ at the tails of the distribution or where we have sparse cluster of data within a given region.

In 2004, Mazio and Taylor proposed an algorithm in which a kernel density classifier is boosted by suitably re-weighting the data. The weight placed on the kernel estimator, is a ratio of a log function in which the denominator is a leave-one-out estimate of the density function. A theoretical explanation is also given by Mazio and Taylor (2004) to show how boosting is a bias reduction technique i.e a reduction of the bias term in the expression for the asymptotic mean integrated squared error (AMISE).

In section 2, we discussed the method boosting in KDE using higher-order hybrid Gaussian kernel. Section 3 is devoted to discussion of data, implementation of algorithm while section 4 concludes this article.

Methods

The leave-one-out estimator of Mazio and Taylor (2004) in the weight function is replaced by a meshsize estimator due to the time complexity involved in the leave-one-out estimator. In the leave-one-out estimator, (n+(n-1)).n function evaluations of the density is required for each boosting step (n- is the sample size). Thus, we are using a meshsize in its place. The only limitation on this meshsize algorithm is that we must first determine the quantity $1/nh$ so as to know what the meshsize that would be placed on the weight function (Ishiekwene *et al.*, 2008, Ishiekwene & Nwelih, 2011). The need to use a meshsize in place of the leave-one-out lies on the fact that boosting is like the steepest-descent algorithm in unconstrained optimization and thus a good substitute that approximates the leave-one-out estimate of the function (Duffy and Helmbold, 2000; Taha, 1971; Ratsch *et al.*, 2000; Mannor *et al.*, 2001: Hazelton & Turlach,2007).

The new meshsize algorithm is stated as:

Meshsize boosting Algorithm for Higher-order Hybrid Gaussian Kernel

STEP 1: Given { x_i , i =1,2,…n} initialize $W_1(i)$= 1/n

STEP 2: Select h(the smoothing parameter)

STEP 3: For m=1,2,…,M

(i) Get

$$\hat{f}_m(x) = \sum_{1}^{n} \frac{W_m(i)}{h} \frac{1}{2\sqrt{2\pi}} \left[1-(t-t_i)^2\right]\left[3-(t-t_i)^2\right] \exp - \frac{(t-t_i)^2}{2}$$

(ii) Update

$$W_{m+1}(i) = W_m(i) + \text{mesh}$$

STEP 4: Provide output

$$\prod_{1}^{M} \hat{f}_m(x)$$

and normalized to integrate to unity.

We can see that the weight function uses a meshsize instead of the leave-one-out log ratio function of Mazio and Taylor (2004). The kernel function used is the higher-order Gaussian kernel unlike the fixed used in Ishiekwene *et al* 2008. The idea of higher-order kernels via bias reduction dates back to Parzen (1962) and Bartlett (1963). Schucany and Summers (1997) also applied the generalized jackknife to bias reduction in kernel density estimation and showed that it is equivalent to using higher-order kernels (Birke, 2009). Jones and Foster (1993) discussed several methods for constructing higher-order kernels. The modified work of Jones and Foster (1993) leads to the proposed higher-order hybrid Gaussian kernel (Afere, 2010) used in this paper. The numerical verification of this algorithm would be seen in the discussion.

Discussion

This section presents three sets of data that are used to illustrate our algorithm and BASIC programming language is used. Table 1 is a sample of size forty and is the lifespan of car batteries in years. Table 2 is a sample of size sixty-four and is the number of written words without mistakes in every 100 words by a set of students in a written essay. Table 3 is the scar length of patients randomly selected in millimeters (Ishiekwene and Afere, 2001; Ishiekwene and Osemwenkhae, 2006).

The results are shown in figures 3.1a – 3.3b. Figure 3.1a is the graph for Data 1 showing the bias reduction, Figure 3.1b for Table 1 showing the MISE. Figure 3.2a is the graph for Table 2 showing the bias reduction, Figure 3.2b for Table 2 showing the MISE. Figure 3.3a is the graph for Table 1 showing the bias reduction, Figure 3.3b for Table 3 showing the MISE. In all three data sets used in this paper, we can clearly see the bias reduction which in turn translates to a reduction in the MISE. Table 3.1 shows the various window widths, $bias^2$, variance and the MISE for all three data sets.

Conclusion

Higher-order hybrid Gaussian kernel has been used in place of the classical fixed kernel in boosting in kernel density estimation and the results compared. The charts- figs. 3.1a – 3.3b and table 3.1 clearly reveals that the higher-order hybrid Gaussian kernel method does better than the classical fixed kernel method in kernel density estimation. It is therefore recommended for use in place of the classical fixed kernel method in boosting in KDE having exhibited the qualities of bias reduction which translates to a reduction in the MISE(ie $bias^2$ + var).

Table 3.1: Showing the various higher-order hybrid window widths, bias, variance and MISE for three data sets

m	n = 40				n = 64				n = 110			
	h_{opt}^{m}	$\int \text{Bias}^2 dx$	$\int \text{Var} dx$	MISE	h_{opt}^{m}	$\int \text{Bias}^2 dx$	$\int \text{Var} dx$	MISE	h_{opt}^{m}	$\int \text{Bias}^2 dx$	$\int \text{Var} dx$	MISE
2	0.304267	0.00366687	0.029335	0.0330019	3.45847	0.000201626	0.00161301	0.00181463	0.139036	0.00291803	0.0233442	0.0262622
4	0.351465	0.0021163	0.0253956	0.0275119	4.05966	0.000114511	0.00137414	0.00148865	0.166255	0.00162686	0.0195223	0.0211492
6	0.388507	0.00143589	0.0229743	0.0244102	4.52586	0.000077037	0.00123259	0.00130963	0.187173	0.00108379	0.0173406	0.0184244
8	0.418113	0.00106737	0.0213475	0.0224148	4.89647	0.0000569648	0.0011393	0.00119626	0.203733	0.000796555	0.0159311	0.0167277
10	0.442283	0.000840869	0.0201809	0.0210217	5.19811	0.000044716	0.00107319	0.0011179	0.217178	0.000622702	0.0149449	0.0155676
12	0.4624	0.000689389	0.0193029	0.0199923	5.44865	0.0000365656	0.00102384	0.0010604	0.228327	0.000507683	0.0142151	0.0147228
14	0.479423	0.000581797	0.0186175	0.0191993	5.66035	0.0000307983	0.000985547	0.00101635	0.237735	0.000426642	0.0136525	0.0140792
16	0.494032	0.00050186	0.0180669	0.0185688	5.84182	0.0000265258	0.000954931	0.000981456	0.245793	0.000366805	0.013205	0.0135718
18	0.506723	0.000440362	0.0176145	0.0180548	5.99931	0.0000232465	0.000929862	0.000953109	0.25278	0.000320999	0.01284	0.013161
20	0.517861	0.000391719	0.0172356	0.0176273	6.13743	0.0000206576	0.000908936	0.000929594	0.258904	0.000284916	0.0125363	0.0128212

References

Afere, B.A (2010) – A family of multivariate higher-order hybrid polynomial kernels in kernel density estimation. Ph.D Thesis, University of Benin, Benin City Nigeria.

Bartlett, M.S. (1963) – Statistical estimation of density functions. Sankhyā series A, **25**, pg 245 – 254.

Birke, Melanie (2009) – Shape constrained KDE. *Journal of Statistical planning & inference, vol* ***139,*** issue 8, August 2009, pg 2851 – 2862.

Duffy, N. and Hemlbold, D. (2000) – Potential bosters? Advances in *Neural info. Proc. Sys.* ***12****,* 258 – 264.

Freund, Y. (1995). Boosting a weak learning algorithm by majority. Info. Comp. **121**,pg 256 - 285.

Hazelton, M.L & Turlach,B.A (2007) – Reweighted KDE. *Computational Statistics & Data Analysis,Vol* ***51***,issue 6,March 2007,pg 3057 – 3069.

Ishiekwene, C. C. and Osemwenkhae, J. E. (2006). – *A comparison of fourth order window width selectors in Kernel Density Estimation* (A Univarate case), ABACUS, **33**; 14 - 20.

Ishiekwene, C.C and Afere, B.A.E. (2001). – Higher Order Window Width Selectors for Empirical Data. *Journal of Nigerian Statistical Association (J.N.S.A),* ***14****,* 69 – 82.

Ishiekwene, C.C; Ogbonmwan, S.M and Osemwenkhae, J.E. (2008). – A Meshsize Boosting Algorithm in Kernel Density Estimation. *Journal of Science & Technology,* Ghana. Vol **28** # 2,pg 69 – 74.

Ishiekwene,C.C. and Nwelih, E. (2011) – Adaptive Kernel in Meshsize Boosting Algorithm in KDE. *African Research Review,5*#1,pg 438-446.

Jones, M.C. and Foster, P.J. (1993) – Generalized Jackknifing and higher-order kernels. *Journal of Nonparametric Statistics. 3,* pg 81 – 94.

Mannor, S.; Meir, R. and Mendelson, S. (2001) – On consistency of boosting algorithms; submitted to Adv. *In Neural info. Proc. Sys.*

Mazio, D. M. and Taylor, C.C. – Boosting Kernel Density estimates: A bias reduction technique? *Biometrika **91***, pg 226 - 233.

Parzen, E. (1962). On the estimation of a probability function and the nodes. Annals of Mathematical Statistics. **33**,pg 1065 – 1076.

Ratsch, G. Mika, S., Scholkopf, B. and Muller, K. K. (2000) – Construction boosting algorithms from SVM's: An application to-one-class classification. *GMD tech report no. 1*.

Schapire, R. (1990). The strength of weak learnability. *Machine learn. **5***,pg 313 – 321.

Schapire, R. and Singer, Y. (1999). – Improved boosting algorithm using confidence rated prediction. *Machine learn. **37***,pg 297 – 336.

Schucany, W.R. and Sommers, J.P. (1977) – Improvement of kernel-type density estimators. *Journal of American Statistical Assoc. **72***, pg 420 -423.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Taha, H. A. (1971). *Operations Research: An Introduction*. Prentice Hall, New Jersey.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. Chapman and Hall/CRC, London.

Table 1

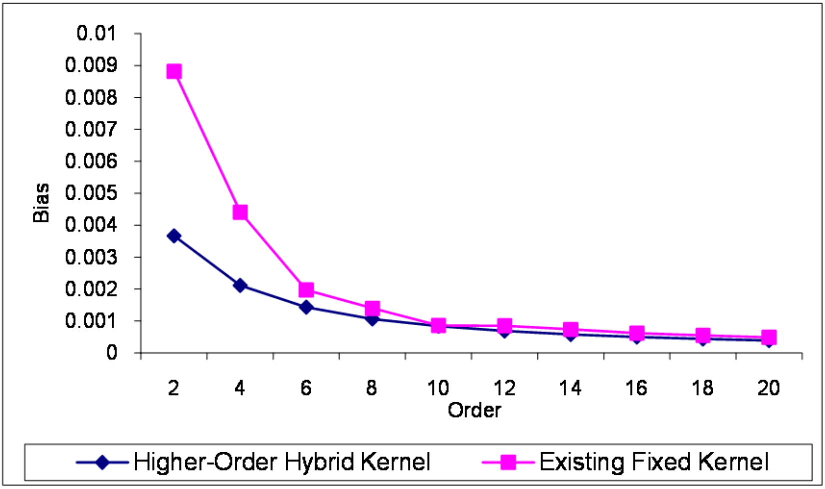
2.2	3.4	2.5	3.3	4.7
4.1	1.6	4.3	3.1	3.8
3.5	3.1	3.4	3.7	3.2
4.5	3.3	3.6	4.4	2.6
3.2	3.8	2.9	3.2	3.9
3.7	3.1	3.3	4.1	3.0
3.0	4.7	3.9	1.9	4.2
2.6	3.7	3.1	3.4	3.5

Table 2

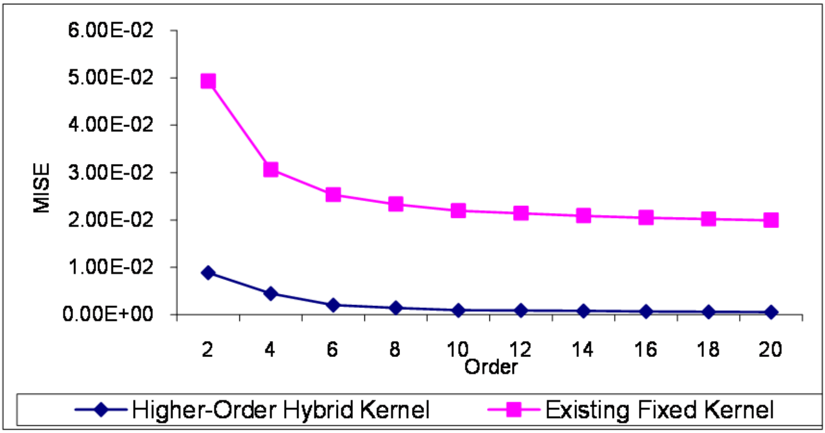
88	69	70	74	70	86	76	74
58	84	68	79	75	83	93	78
92	85	69	67	81	79	97	83
77	78	84	68	80	69	87	69
81	79	88	96	77	83	75	91
86	72	89	90	79	73	83	88
90	86	82	66	80	75	81	82
67	94	75	69	91	85	76	80

Table 3

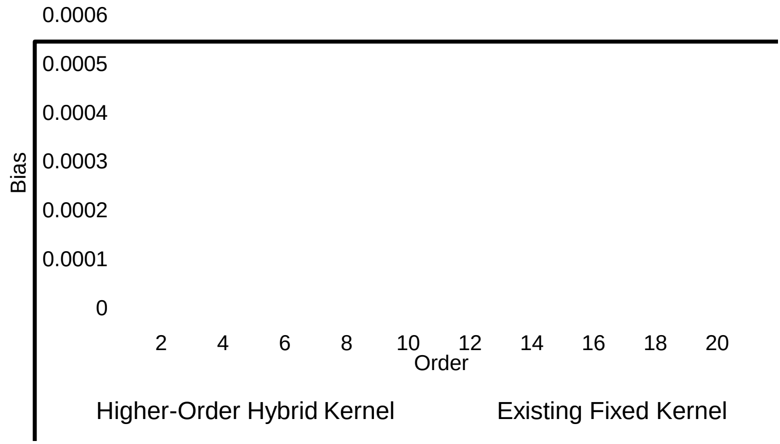
1.2	2.0	1.6	2.4	2.0
1.4	2.4	2.4	2.0	2.4
2.6	1.8	2.0	2.6	1.9
2.0	1.6	1.4	1.3	2.0
1.4	1.64	1.6	1.7	2.4
1.7	1.3	1.8	1.6	2.0
1.6	2.0	1.2	1.5	1.98
1.5	1.9	2.0	1.9	2.2
1.48	1.4	2.2	2.4	1.6
1.6	2.0	1.8	2.1	2.4
2.2	1.4	1.9	2.3	2.6
1.35	1.7	2.0	1.8	2.0
1.35	1.9	2.3	1.4	1.6
1.2	1.6	1.4	1.9	1.7
1.6	2.0	1.8	2.0	1.9
1.2	2.4	1.64	1.3	2.2
1.6	1.8	2.0	1.9	1.86
1.2	1.6	2.3	1.42	1.4
2.0	1.64	1.2	1.47	1.9
1.4	1.3	1.3	1.4	1.7
1.7	1.4	1.9	1.9	1.6
1.6	2.4	2.0	2.0	2.3



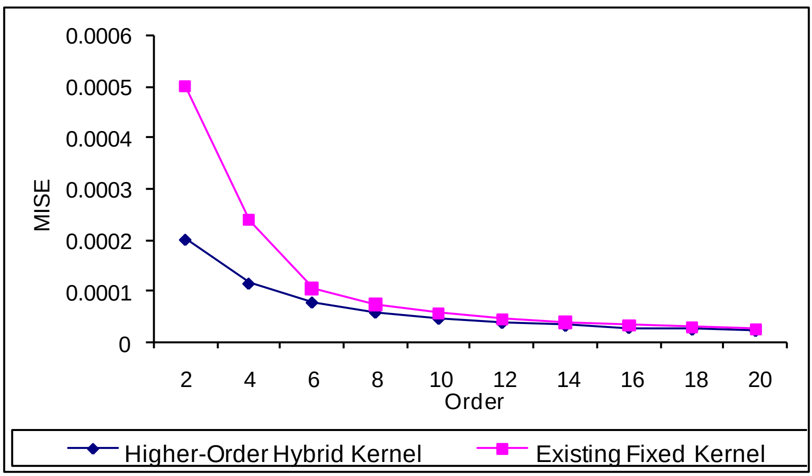
Graph showing the Bias for Table 1



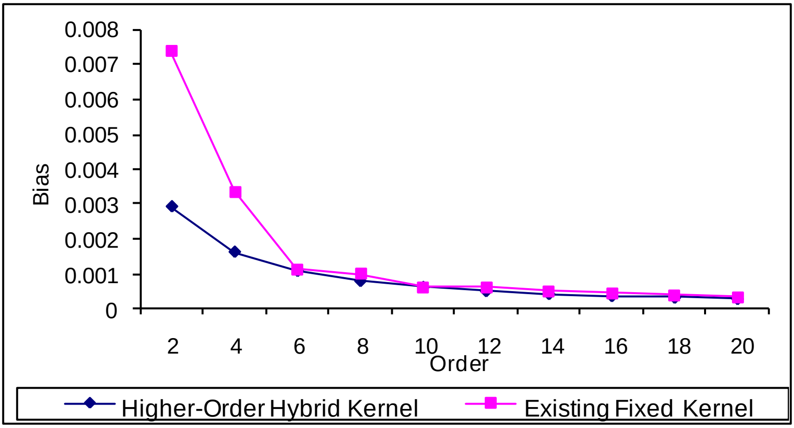
Graph showing the MISE for Table 1



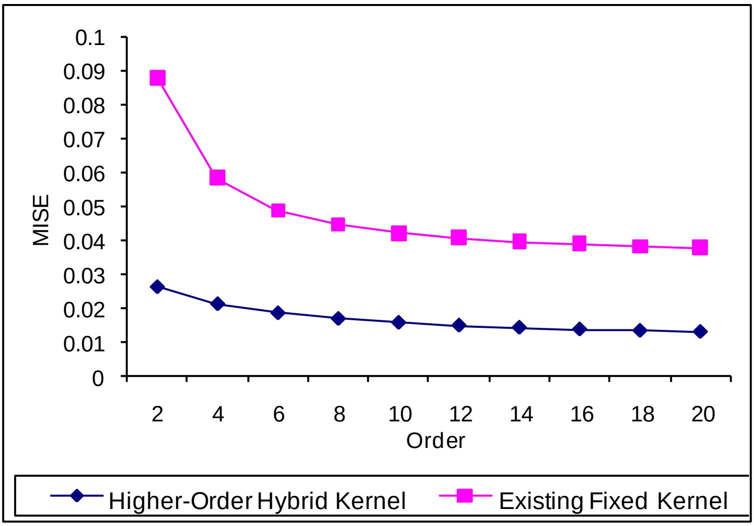
Graph showing the Bias for Table 2



Graph showing the MISE for Table 2



Graph showing the Bias for Table 3



Graph showing the MISE for Table 3