DOWNLOAD PDF NONPARAMETRIC ESTIMATION OF PROBABILITY DENSITIES AND REGRESSION CURVES

Chapter 1 : Kernel density estimation - Wikipedia

of Probability Densities and Regression Curves Translated by NONPARAMETRIC ESTIMATION OF THE REGRESSION CURVE AND COMPONENTS OF A CONVOLUTION I. Some.

Testing serial independence using the sample distribution function by Miguel A. Delgado - Journal of Time Series Analysis , " Mises statistic based on the difference between the joint sample distribution and the product of the marginals. Exact critical values can be approximated from the asymptotic null Exact critical values can be approximated from the asymptotic null distribution or by resampling, randomly permuting the original series. A Monte Carlo experiment illustrates the test performance with small and moderate sample sizes. The paper also includes an application, testing the random walk hypothesis of exchange rate returns for several currencies. Key Words Serial independence test; Multivariate sample distribution; Hoeffding-Blum-Kiefer-Rosenblatt empirical process; Random permutation test; Ergodic alternatives. Show Context Citation Context The authors consider the estimation of a set S IR d from a random sample of n points. They obtain the convergence rate for the probability of false alarm and show that the smoothing parameter n can be used to incorporate some prior information on the shape of S. They suggest two general methods for selecting n and illustrate them with a simulation study and a real data example. Efficient regression in metric spaces via approximate Lipschitz extension by Lee-ad Gottlieb, Aryeh Kontorovich, Robert Krauthgamer , " We present a framework for performing efficient regression in general metric spaces. Roughly speaking, our regressor predicts the value at a new point by computing a Lipschitz extension â€" the smoothest function consistent with the observed data â€" while performing an optimized structural ri Roughly speaking, our regressor predicts the value at a new point by computing a Lipschitz extension â€" the smoothest function consistent with the observed data â€" while performing an optimized structural risk minimization to avoid overfitting. The offline learning and online inference stages can be solved by convex programming, but this naive approach has runtime complexity O n 3 , which is prohibitive for large datasets. We make dual use of the doubling dimension: Our resulting regressor is both asymptotically strongly consistent and comes with finite-sample risk bounds, while making minimal structural and noise assumptions. Rao - in Multisensor Fusion, A. Hyder editor , " We consider a multiple sensor system such that for each sensor the outputs are related to the actual feature values according to a certain probability distribution. We present an overview of informational and computational aspects of a fuser that is required to combine the sensor outputs to more acc We present an overview of informational and computational aspects of a fuser that is required to combine the sensor outputs to more accurately predict the feature, when the sensor distributions are unknown but iid measurements are given. Our focus is on methods to compute a fuser with probabilistic guarantees in terms of distribution-free performance bounds based on a finite sample. We first discuss a number of methods based on the empirical risk minimization approach. These methods yield a fuser which is guaranteed, with a high probability, to be close to an optimal fuser computable only under a complete knowledge of sensor distributions. Then we describe the isolation fusers that are guaranteed to perform at least as good as the best sensor, and the projective fusers that are guaranteed to perform at least as good as th This property is the key to efficient computation of the estimate Rao and Protopopescu, This estimator was found to be very effective in a number of applications involving nonlinear regression estimation. The typical results about the performance of this estimator are in terms of asym Density-based distance metrics have applications in semi-supervised learning, nonlinear interpolation and clustering. We consider density-based metrics induced by Riemannian manifold structures and estimate them using kernel density estimators for the underlying data distribution. We lower bound the We lower bound the rate of convergence of these plug-in pathlength estimates and hence of the metric, as the sample size increases. We present an upper bound on the rate of convergence of all estimators of the metric. We also show that the metric can be consistently computed using the shortest path algorithm on a suitably constructed graph on the data samples and lower bound the convergence rate of the computation error. We present experiments illustrating the use of the metrics for semisupervised classification and non-linear interpolation. We use the following two lemma about well known, e. Lemma 1 Bias of the kernel density estimator. Accurate and fast estimation of probability density functions is crucial for satisfactory computational performance in many scientific problems. When the type of density is known a priori, then the problem becomes statistical estimation of parameters from the observed values. In the nonparametric ca In the nonparametric case, usual estimators make use of kernel functions. We propose a sequence of special weight functions for the nonparametric estimation of f which requires almost linear time: We derive conditions for convergence under a number of metrics, which turn out to be similar to those Extensive descriptions of various approaches to non-parametric estimation along with a comprehensive bibliography can be found in relatively recent books by B. Silverman [23], and E. Results of the experimental comparison of some widely used methods appear in [7, 24]. Along with applications in discriminant analysis and cluster analysis, an important application of non-parametri Key words and phrases: The LIL for L2 functionals of kernel density estimators 1. We investigate the asymptotic behavior of the Nadaraya-Watson esti-mator for the estimation of the regression function in a semiparametric regression model. On the one hand, we make use of the recursive version of the sliced inverse regression method for the estimation of the unknown param On the one hand, we make use of the recursive version of the sliced inverse regression method for the estimation of the unknown parameter of the model. On the other hand, we implement a recursive Nadaraya-Watson procedure for the estimation of the regression function which takes into account the previous esti-mation of the parameter of the semiparametric regression model. We establish the almost sure convergence as well as the asymptotic normality for our Nadaraya-Watson estimator. We also illustrate our semiparametric estimation procedure on simulated data. One can find a wide range of literature on nonparametric estimation of a regression function. Representations for integral functionals of kernel density estimators by David M. Mason - Austrian J. Statist , " We establish a representation as a sum of independent random variables, plus a remainder term, for estimators of integral functionals of the density function, which have a certain simple structure. From this represen-tation we derive a central limit theorem, a law of large numbers and a la From this represen-tation we derive a central limit theorem, a law of large numbers and a law of the iterated logarithm. Some motivating special cases are T1 f: Rao , " The problem is to estimate a fusion rule f: Let f minimize I:

Chapter 2 : Density Estimation Using Regression

Nonparametric Estimation of Probability Densities and Regression Curves (Mathematics and its Applications) th Edition.

Local linear estimation of irregular curves with applications. Local linear fitting and improved estimation near peaks. The Canadian Journal of Statistics 37 â€” Curve fitting under jump and peak irregularities using local linear regression. Communications in Statisticsâ€"Theory and Methods, to appear. Smoothing and preservation of irregularities using local linear fitting. Applications of Mathematics 53 â€" Bandwidth selection for change point estimation in nonparametric regression. MR Digital Object Identifier: Edge-preserving image denoising and estimation of discontinuous surfaces. Jump-preserving regression and smoothing using local linear fitting: The Annals of the Institute of Statistical Mathematics 59 â€" Some new methods for wavelet density estimation. Sankhya Series A 63 â€" Local bandwidth choice in kernel regression estimation. Journal of Computational and Graphical Statistics 6 35â€" Generalized jackknifing and higher order kernels. Journal of Nonparametric Statistics 3 81â€" Adaptive variable location kernel density estimators with good-performance at boundaries. Journal of Nonparametric Statistics 15 61â€" Nonparametric estimations of discontinuous curves and surfaces. Kernel estimators for probability densities with discontinuities. Transformations to reduce boundary bias in kernel density estimation. Exact mean integrated squared error. The Annals of Statistics 20 â€"

Chapter 3 : Nonparametric regression - Wikipedia

Nonparametric Estimation of Probability Densities and Regression Curves. Authors Nonparametric Estimation of Regression Curves and Components of a Convolution.

Gaussian process regression In Gaussian process regression, also known as Kriging, a Gaussian prior is assumed for the regression curve. The errors are assumed to have a multivariate normal distribution and the regression curve is estimated by its posterior mode. The Gaussian prior may depend on unknown hyperparameters, which are usually estimated via empirical Bayes. Smoothing splines have an interpretation as the posterior mode of a Gaussian process regression. Kernel regression Example of a curve red line fit to a small data set black points with nonparametric regression using a Gaussian kernel smoother. The pink shaded area illustrates the kernel function applied to obtain an estimate of y for a given value of x. The kernel function defines the weight given to each data point in producing the estimate for a target point. Nonparametric multiplicative regression[edit] Two kinds of kernels used with kernel smoothers for nonparametric regression. Use of Gaussian kernels for nonparametric multiplicative regression with two predictors. The weights from the kernel function for each predictor are multiplied to obtain a weight for a given data point in estimating a response variable dependent variable at a target point in the predictor space. Two commonly used forms of a local model used in nonparametric regression, contrasted with a simple linear model. Nonparametric multiplicative regression NPMR is a form of nonparametric regression based on multiplicative kernel estimation. Like other regression methods, the goal is to estimate a response dependent variable based on one or more predictors independent variables. NPMR can be a good choice for a regression method if the following are true: The shape of the response surface is unknown. The predictors are likely to interact in producing the response; in other words, the shape of the response to one predictor is likely to depend on other predictors. This is a smoothing technique that can be cross-validated and applied in a predictive way. Organismal response to environment tends to be nonlinear and have complex interactions among predictors. NPMR allows you to model automatically the complex interactions among predictors in much the same way that organisms integrate the numerous factors affecting their performance. For example, assume that a plant needs a certain range of moisture in a particular temperature range. If either temperature or moisture fall outside the tolerance of the organism, then the organism dies. If it is too hot, then no amount of moisture can compensate to result in survival of the plant. Mathematically this works with NPMR because the product of the weights for the target point is zero or near zero if any of the weights for individual predictors moisture or temperature are zero or near zero. Note further that in this simple example, the second condition listed above is probably true: Optimizing the selection of predictors and their smoothing parameters in a multiplicative model is computationally intensive. With a large pool of predictors, the computer must search through a huge number of potential models in search for the best model. The best model has the best fit, subject to overfitting constraints or penalties see below. By "local model" we mean the way that data points near a target point in the predictor space are combined to produce an estimate for the target point. The most common choices for the local models are the local mean estimator, a local linear estimator, or a local logistic estimator. In each case the weights can be extended multiplicatively to multiple dimensions. In words, the estimate of the response is a local estimate for example a local mean of the observed values, each value weighted by its proximity to the target point in the predictor space, the weights being the product of weights for individual predictors. The model allows interactions, because weights for individual predictors are combined by multiplication rather than addition. Overfitting controls Understanding and using these controls on overfitting is essential to effective modeling with nonparametric regression. Nonparametric regression models can become overfit either by including too many predictors or by using small smoothing parameters also known as bandwidth or tolerance. This can make a big difference with special problems, such as small data sets or clumped distributions along predictor variables. Instead, one can control overfitting by setting a minimum average neighborhood size, minimum data: Nonparametric regression models sometimes use an AIC based on the "effective number of parameters". If, however, you use leave-one-out cross validation in the model fitting phase, the trace of the smoothing matrix is always zero, corresponding to zero parameters for the AIC. Thus, NPMR with cross-validation in the model fitting phase already penalizes the measure of fit, such that the error rate of the training data set is expected to approximate the error rate in a validation data set. In other words, the training error rate approximates the prediction extra-sample error rate. Related techniques NPMR is essentially a smoothing technique that can be cross-validated and applied in a predictive way. Many other smoothing techniques are well known, for example smoothing splines and wavelets. The optimal choice of a smoothing method depends on the specific application. Nonparametric regression models always fits for larger data. Decision tree learning Decision tree learning algorithms can be applied to learn to predict a dependent variable from data.

Chapter 4 : CiteSeerX â€" Citation Query Nonparametric Estimation of Probability Densities and Regressio

Nonparametric Estimation of the Regression Curve and Components of a Convolution.- 1. Some Asymptotic Properties of Nonparametric Estimators of Regression Curves

Chapter 5 : Desmet, Gijbels, Lambert: Estimation of irregular probability densities

Get this from a library! Nonparametric Estimation of Probability Densities and Regression Curves. [E A Nadaraya] -- 'Et moi , si. j'avail su comment en revenir.

Chapter 6 : Modal regression using kernel density estimation: A review - WIREs Computational Statistics

Efficient Non-parametric Estimation of Probability Density Functions by Ã–mer Egecioglu, Ashok Srinivasan - SIAM J. of Scientific Computing, Accurate and fast estimation of probability density functions is crucial for satisfactory computational performance in many scientific problems.