Master Thesis Proposal: Robust Kernel Density Estimation for Naive Bayes Classifier

Boli Bi Iritié A-D

Sep 27, 2022

1. What’s Robust Kernel Density Estimation

Kernel Density Estimation is a Nonparametric approach for calculating probabilities for new points using a dataset.

A kernel is a mathematical function that calculates the likelihood of a random variable at a given value. The kernel effectively smooths or interpolates the probabilities across the range of possible outcomes for a random variable, ensuring that the total of probabilities equals one, as is required by well-behaved probabilities.

The kernel function weights the contribution of observations from a data sample based on their relationship or distance to a given query sample for which the probability is requested.

The scope, or window of observations, from the data sample that contributes to estimating the probability for a given sample is controlled by a parameter called the **smoothing parameter** or **the bandwidth**. As a result, kernel density estimation is sometimes known as a Parzen-Rosenblatt window, or simply a Parzen window, after the method's creators.

* **Smoothing Parameter (bandwidth)**: Parameter that controls the number of samples or window of samples used to estimate the probability for a new point.

A big window may produce a coarse density with few features, whereas a tiny window may include too much detail and be too smooth or general to cover new or unseen cases adequately. Varying functions, commonly referred to as fundamental functions, such as uniform normal, etc., can be used to shape the contribution of samples within the window, with different impacts on the smoothness of the final density function.

* **Basis Function (kernel)**: The function chosen used to control the contribution of samples in the dataset toward estimating the probability of a new point.

As a result, experimenting with different window sizes and contribution functions and comparing the results to data histograms may be beneficial.

1. What’s Naïve Bayes Classifier
2. Robust Kernel Density Estimation for Naïve Bayes Classifier

**References**

1. A Gentle Introduction to Probability Density Estimation by Jason Brownlee, September 25, 2019 [online] available: <https://machinelearningmastery.com/probability-density-estimation/>
2. Probability density estimation in higher dimensions by David W. Scott and James R. Thompson, [online] available: <https://www.researchgate.net/profile/David-Scott-57/publication/247260738_Probability_density_estimation_in_higher_dimension/links/53f8b0100cf24ddba7db47a1/Probability-density-estimation-in-higher-dimension.pdf>
3. Robust kernel density estimation by JooSeuk Kim and Clayton Scott, 2008 [online] available: IEEE Xplore
4. <https://en.wikipedia.org/wiki/Naive_Bayes_classifier>