Lehrstuhl für STEUERUNGS- UND REGELUNGSTECHNIK Technische Universität München Prof. Dongheui Lee	MACHINE LEARNING IN ROBOTICS Exercises 8: Non-parametric Density Estimation

Exercise 1

In density estimation with Parzen windows, we distribute hypercubes of fixed size in the dataspace and the number of samples involved in these hypercubes is estimated based on the given data distribution. Assume that an 1D dataset is observed and the number of samples per hypercube is given by $K = \sum_{i=1}^{n} k(\frac{x-x^{(i)}}{h})$ where

$$k(u) = \begin{cases} 1 & |u| \leq 1/2 \\ 0 & otherwise \end{cases}$$

a) Prove that the pdf of the data distribution is given by $p(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h} \cdot k(\frac{x-x^{(i)}}{h})$.

b) Analyze the effect that the window width has on the estimation of the function $p(x)$.

Solution

(a)The general formula giving the pdf of the data distribution is given by:

$$p(x) = \frac{K}{nV} = \frac{\sum_{i=1}^{n} k(\frac{x-x^{(i)}}{h})}{nh}$$

(b)

$$p_n(x) = \frac{1}{nh}\sum_{i=1}^{n} k(\frac{x-x^{(i)}}{h})$$

h affects the amplitude of the density function and its width.

- if h large, the amplitude decreases and the density function $p(x)$ becomes smoother and maybe oversmoothed.
- if h small, $p(x)$ becomes high and this increase occurs where x is very near to the $x^{(i)}$. Thus, $p(x)$ approaches a Dirac delta function centered at $x^{(i)}$ and $p(x)$ arises from a superposition of sharp pulses centered too close to the sample points.

Exercise 2

Consider a histogram-like density model in which the space x is divided into fixed regions for which the density $p(x)$ takes the constant value h_k over the k^{th} region, and that the volume of region k is denoted Δ_k . Suppose we have a set of n observations of x such that n_k of these observations fall in region k. Using a Lagrange multiplier to enforce the normalization constraint on the density, derive an expression for the maximum likelihood estimator for the h_k.

Solution

The value of the density $p(x)$ at a point $x^{(i)}$ is given by $h_{j(i)}$, where the notation $j(i)$ denotes that data point $x^{(i)}$ falls within region j. Thus the log-likelihood function takes the form

$l = \sum_{i=1}^{n} \ln p(x^{(i)}) = \sum_{i=1}^{n} \ln h_{j(i)}$

We now need to take account of the constraint that $p(x)$ must integrate to unity. Since $p(x)$ has the constant value h_r over region r, which has volume Δ_r , the normalization constraint becomes $\sum_r h_r \Delta_r = 1$. Introducing a Lagrange multiplier λ we then maximize the function

$\sum_{i=1}^{n} \ln h_{j(i)} + \lambda \left(\sum_r h_r \Delta_r - 1 \right)$

Taking partial derivative with respect to h_k and evaluating it to zero gives:

$0 = \frac{n_k}{h_k} + \lambda \Delta_k$

$0 = n_k + \lambda \Delta_k h_k$

where n_k denotes the total number of data points falling within region k.

Now making use of the normalization constraint,

$0 = \sum_r n_r + \lambda (\sum_r h_r \Delta_r)$

$\Rightarrow \lambda = -n$

Eliminating λ then gives our final result for the maximum likelihood solution for h_k in the form

$h_k = \frac{n_k}{n} \frac{1}{\Delta_k}$

Note that, for equal sized bins $\Delta_k = \Delta$ we obtain a bin height h_k which is proportional to the fraction of points falling within that bin, as expected.

Exercise 3

Show that the K-nearest-neighbour density model defines an improper distribution whose integral over all space is divergent.

Solution

Lets consider a 1-dimensional problem with fixed K. The density estimate is given by:

$p(x) = \frac{K}{nV}$

where n is the total number of data points and V is the width of the bin. Now if we place the bin at the right most point x_r, then with width Δ_r it encloses k points. Now if we move by an amount δ to the right of x_r, the bin width will increase by 2δ and the density estimate of that point will be:

$p(x_r + \delta) = \frac{K}{n(\Delta_r + 2\delta)}$

To be a proper density estimate, a probability density function should integrate to 1. Now for testing the convergence, if we integrate δ from $0 \to \infty$ we get

$$\int_{\delta=0}^{\infty} p(x_r + \delta) d\delta = \int_{\delta=0}^{\infty} \frac{K}{n(\Delta_r + 2\delta)} d\delta = \frac{K}{2n} \ln \left| n(\Delta_r + 2\delta) \right| \Big|_{\delta=0}^{\infty} = \infty$$

$\left(\text{Integral of the above form is calculated as: } \int \frac{c}{ax+b} dx = \frac{c}{a} \ln |ax+b| + C \right)$

and thus the integral is divergent and the density model is an improper distribution.