CZ4041/CE4041: Machine Learning

Lecture 11: Density Estimation

Question 1

- Suppose a dataset of four 3-dimensional instances is shown below. Estimate the sample mean and covariance matrix (unbiased).

Data matrix

X

	X_1	X_2	X_3
P1	3	5	−1
P2	−1	8	3
P3	2	−4	−4
P4	0	−1	−6

$$\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i$$

$$\hat{\boldsymbol{\mu}} = \frac{1}{4}\sum_{i=1}^{4} Pi$$

$\hat{\boldsymbol{\mu}}$

X_1	X_2	X_3
1	2	−2

Question 1 (cont.)

Centered data matrix

$$\widetilde{\mathbf{\Sigma}} = \frac{1}{N-1}\sum_{i=1}^{N} \underbrace{\boxed{(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})}}_{}^{\widetilde{\boldsymbol{x}}_i}(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})^T = \frac{1}{N-1}\sum_{i=1}^{N} \widetilde{\boldsymbol{x}}_i\,\widetilde{\boldsymbol{x}}_i^T = \frac{1}{N-1}\widetilde{\mathbf{X}}^T\boxed{\widetilde{\mathbf{X}}}$$

$\mathbf{X}$

X_1	X_2	X_3
3	5	−1
−1	8	3
2	−4	−4
0	−1	−6

X_1	X_2	X_3
1	2	−2

$\widehat{\boldsymbol{\mu}}$

$\widetilde{\mathbf{X}}$

X_1	X_2	X_3
2	3	1
−2	6	5
1	−6	−2
−1	−3	−4

$$\widetilde{\mathbf{\Sigma}} = \frac{1}{4-1}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}} =$$

3.33	−3	−2
−3	30	19
−2	19	15.33

Question 2

- Suppose a dataset of five 1-dimensional instances is shown in Table 2. Use histogram estimator with an origin of 0 and a width of 3, naive estimator with a width of 3, and 3-NN estimator to estimate the density function $\hat{p}(\boldsymbol{x})$ and compute the value of $\hat{p}(2.6)$ at 2.6, respectively.

Table 2: Data set for Question 2.

P1	P2	P3	P4	P5
1.2	2	10	-5.9	3.5

Histogram Estimator

- Simply partition $\boldsymbol{x}$ into distinct bins of a fixed width Δ
- Count the number N_t of training data points falling into bin t
- Turn this count into a normalized probability density via dividing by the total number of observed data points N and by the width Δ of the bins:

$$p_t = \frac{N_t}{N\Delta}$$

- The model for the density $p(\boldsymbol{x})$ is constant over the width of each bin: find the bin where $\boldsymbol{x}$ is in (e.g., bin t), then

$$\hat{p}(\boldsymbol{x}) = \frac{\#\{\boldsymbol{x}_i \mid \boldsymbol{x}_i \text{ in the same bin as } \boldsymbol{x}\}}{N\Delta} = p_t$$

Histogram Estimator (cont.)

Training data

P_1	P_2	P_3	P_4	P_5
1.2	2	10	−5.9	3.5

Query $\boldsymbol{x}$ $\boxed{2.6}$

$\boxed{2.6}$

Global intervals:	$(-6,-3]$	$(-3,0]$	$(0,3]$	$(3,6]$	$(6,9]$	$(9,12]$
Counts:	1	0	2	1	0	1

$$\hat{p}(2.6) = \frac{\#\{P_i \mid P_i \text{ in the same bin as } 2.6\}}{N\Delta} = \frac{2}{5 \times 3} = 0.13$$

Naïve Estimator: An Alterative

- In Histogram Estimator, besides Δ, we have to choose an origin x_0 as well, the bins are the intervals defined as

$$(\boldsymbol{x}_0 + \boxed{m}\Delta, \boldsymbol{x}_0 + (m+1)\Delta]$$

0, positive or negative integers

- The Naïve Estimator does not need to set an origin

$$\hat{p}(\boldsymbol{x}) = \frac{\#\{\boldsymbol{x}_i \mid \boldsymbol{x}_i \text{ in the same bin as } \boldsymbol{x}\}}{N\Delta}$$

Given $\boldsymbol{x}$, use $\boldsymbol{x}$ as a center to create a bin with a length of Δ

$$\left[\xleftrightarrow{\frac{\Delta}{2}} \boldsymbol{x} \xleftrightarrow{\frac{\Delta}{2}}\right]$$

$$\hat{p}(\boldsymbol{x}) = \frac{\#\left\{\boldsymbol{x}_i \mid \boxed{\boldsymbol{x} - \frac{\Delta}{2} < \boldsymbol{x}_i \le \boldsymbol{x} + \frac{\Delta}{2}}\right\}}{N\Delta}$$

Naïve Estimator (cont.)

Training data

P_1	P_2	P_3	P_4	P_5
1.2	2	10	−5.9	3.5

Query $\boldsymbol{x}$ $\boxed{2.6}$

$\boxed{2.6}$

Local interval: $(2.6 - \frac{3}{2}, 2.6 + \frac{3}{2}] \Rightarrow (1.1, 4.1]$

$$\hat{p}(2.6) = \frac{\#\{P_i \mid P_i \text{ in the same bin as } 2.6\}}{N\Delta} = \frac{3}{5 \times 3} = 0.2$$

K-NN Estimator

- The K-NN Estimator *adapts* the amount of smoothing to the *local* density of data, and the degree of smoothing is controlled by K, the number of neighbors

Consider K nearest neighbors of $\boldsymbol{x}$

$$p(\boldsymbol{x}) = \frac{K}{NV_{\boldsymbol{x}}}$$

The volume of the space centered at $\boldsymbol{x}$ that exactly contains K nearest neighbors of $\boldsymbol{x}$

- For the multivariate case

$$\hat{p}(\boldsymbol{x}) = \frac{K}{N\boxed{V_{d_K(\boldsymbol{x})}}}$$

The volume of the d-ball of the radius $d_K(\boldsymbol{x})$ centered at $\boldsymbol{x}$. And $d_K(\boldsymbol{x})$ is the distance of $\boldsymbol{x}$ to the K-th nearest observed instance

- For the univariate case

$$\hat{p}(\boldsymbol{x}) = \frac{K}{N2d_K(\boldsymbol{x})}$$

K-NN Estimator (cont.)

Training data

P_1	P_2	P_3	P_4	P_5
1.2	2	10	−5.9	3.5

Query $\boldsymbol{x}$ $\boxed{2.6}$

Distance between $\boldsymbol{x}$ and the training data

P_1	P_2	P_3	P_4	P_5
1.4	0.6	7.4	8.5	0.9

3NN

3rd NN

$$\hat{p}(\boldsymbol{x}) = \frac{K}{N2d_K(\boldsymbol{x})} \longrightarrow \hat{p}(2.6) = \frac{3}{5 \times 2 \times 1.4} = 0.21$$

Thank you!