

Nearest Neighbour Method

G V V Sharma*

CONTENTS

1	Data	1
2	Nearest Neighbour	1

Abstract—This manual introduces the nearest neighbour method.

1 DATASET

1.1 Generate $S_x = \{\mathbf{m}_{xk}\}_{k=1}^{10}$ from the bivariate Gaussian distribution.

$$\mathcal{N}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, I\right). \quad (1.1)$$

1.2 Generate $S_y = \{\mathbf{m}_{yk}\}_{k=1}^{10}$ from the bivariate Gaussian distribution.

$$\mathcal{N}\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}, I\right). \quad (1.2)$$

1.3 Generate

$$\mathbf{x}_n = \mathcal{N}(\mathbf{m}_{xk}, I/5), n = 1, \dots, 100 \quad (1.3)$$

such that \mathbf{m}_{xk} are picked uniformly from S_x .

1.4 Generate

$$\mathbf{y}_n = \mathcal{N}(\mathbf{m}_{yk}, I/5), n = 1, \dots, 100 \quad (1.4)$$

such that \mathbf{m}_{yk} are picked uniformly from S_y .

1.5 Generate a scatterplot for \mathbf{x}_n in orange colour and \mathbf{y}_n in blue colour.

2 NEAREST NEIGHBOUR

2.1 Find $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$,

$$y_2 = \frac{1}{k} \sum_{i \in N_k(y_1)} x_{2i} \quad (2.1)$$

where

$$N_k(y_1) = \{i : |x_{1i} - y_1| < \epsilon < |x_{1j} - y_1|\}, \quad (2.2)$$

$$|N_k(y_1)| = k \quad (2.3)$$

2.2 Plot $\mathbf{y}_m, 1 \leq m \leq 100$ for $k = 15$.

2.3 Repeat the exercise for $k = 1$.

2.4 Compare with the least squares estimate.

*The author is with the Department of Electrical Engineering, Indian Institute of Technology, Hyderabad 502285 India e-mail: gadepall@iith.ac.in. All content in this manual is released under GNU GPL. Free and open source.