Tutorial 7: Nonparametric Density Estimation

Rui Zhao

rzhao@ee.cuhk.edu.hk

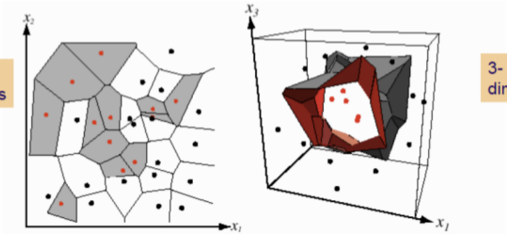
Mar. 5, 2014

Outline

Nearest Neighbor Classifier

- For a test point $\mathbf{x}$, find its nearest sample $\mathbf{x}'$ in the training set and assign it the label associated with $\mathbf{x}'$
- The feature space is partitioned into cells consisting of all points closer to a given training point $\mathbf{x}'$ than to any other training points
- All points in such a cell are labeled by the category of the training point - *Voronoi tesselation* of the space

2-dimensions



3-dimensions

Ex.1: Voronoi cell

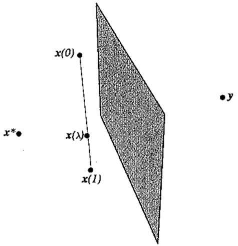
Problem

Prove that the Voronoi cells induced by the single-nearest neighbor algorithm must always be convex. That is, for any two points x_1 and x_2 in a cell, all points on the line linking x_1 and x_2 must also lie in the cell.

Ex.1: Voronoi cell

Solution

Our goal is to show that Voronoi cells induced by the nearest-neighbor algorithm are convex, that is, given any two points in the cell, the line connecting them also lies in the cell. We let x^* be the labeled sample point in the Voronoi cell, and y be any other labeled sample point. A unique hyperplane separates the space into those that are closer to x^* than to y, as shown in the figure. Consider any two points x_0 and x_1 inside the Voronoi cell of x^*; these points are surely on the side of the hyperplane nearest x^*. Now consider the line connecting those points, parameterized as $x(\lambda) = (1 - \lambda)x_0 + \lambda x_1$ for $0 \leq \lambda \leq 1$. Because the half-space defined by the hyperplane is convex, all the points $x(\lambda)$ lie nearer to x^* than to y. This is true for every other sample point y_i. Thus $x(\lambda)$ remains closer to x^* than any other labeled point. By our definition of convexity, we have, then, that the Voronoi cell is convex.

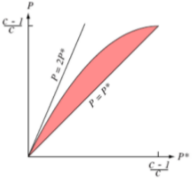


Outline

Error Bound

- Nearest-neighbor classifier is a sub-optimal procedure and leads to an error rate greater than Bayesian error rate
- But with an unlimited number of training samples, it is never worse than twice the Bayes error rate
- Let P^* be the Bayes error rate, $P_n(e)$ be the n-sample error rate, and $P = \lim_{n\to\infty} P_n(e)$, then

$$P^* \le P \le P^* \left(2 - \frac{c}{c-1} P^*\right)$$



Ex.2: Nearest neighbor error rate

Review

Independently drawn labelled samples:

$$(x_1, \theta_1), (x_2, \theta_2), \ldots, (x_n, \theta_n), \tag{1}$$

where θ_j may be any of c states of nature $w_i, i = 1, ..., c$. Given a test (x, θ), its nearest training sample is x_{nn}, we have

$$P(\theta, \theta_{nn}|x, x_{nn}) = P(\theta|x)P(\theta_{nn}|x_{nn}) \tag{2}$$

The nearest-neighbor decision rule: commit an error whenever $\theta \neq \theta_{nn}$

$$P_n(e|x, x_{nn}) = 1 - \sum_{i=1}^{c} P(\theta = w_i, \theta_{nn} = w_i|x, x_{nn}) \tag{3}$$

$$= 1 - \sum_{i=1}^{c} P(\theta = w_i|x)P(\theta_{nn} = w_i|x_{nn}) \tag{4}$$

Ex.2: Nearest neighbor error rate

Review [continue]

$$p(e|x) = \int P(e|x, x_{nn})p(x_{nn}|x)dx_{nn} \tag{5}$$

where $\quad p(x_{nn}|x) \rightarrow \delta(x_{nn} - x), \quad n \rightarrow \infty,$
and $\quad P_n(e|x, x_{nn}) = 1 - \sum_{i=1}^{c} P(\theta = w_i|x)P(\theta_{nn} = w_i|x_{nn})$
Thus,

$$p(e|x) = \lim_{n\rightarrow\infty} P_n(e|x) = \int P_n(e|x, x_{nn})\delta(x_{nn} - x)dx_{nn} \tag{6}$$

$$= 1 - \sum_{i=1}^{c} P^2(\theta = w_i|x) \tag{7}$$

Ex.2: Nearest neighbor error v.s. Bayes error

Problem

It is easy to see that the nearest-neighbor error rate P can equal the Bayes rate P^* if $P^* = 0$ (the best possibility) or if $P^* = \frac{c-1}{c}$ (the worst possibility). One might ask whether or not there are problems for which $P = P^*$ when P is between these extremes.

1. Show that the Bayes rate for the one-dimensional case where $P(w_i) = 1/c$ and

$$P(x|w_i) = \begin{cases} 1 & 0 \leq x \leq \frac{cr}{c-1} \\ 1 & i \leq x \leq i + 1 - \frac{cr}{c-1} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

   is $P^* = r$.

2. Show that for this case that the nearest-neighbor rate is $P = P^*$.

Ex.2: Nearest neighor error v.s. Bayes error

Solution

It is indeed possible to have the nearest-neighbor error rate P equal to the Bayes error rate P^* for non-trivial distribution.

1. Consider uniform priors over c categories, that is, $P(w_i) = 1/c$, and one-dimensional distributions

$$P(x|w_i) = \begin{cases} 1 & 0 \leq x \leq \frac{cr}{c-1} \\ 1 & i \leq x \leq i + 1 - \frac{cr}{c-1} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Note that this automatically impose the restriction

$$0 \leq \frac{cr}{c-1} \leq 1 \tag{10}$$

Ex.2: Nearest neighor error v.s. Bayes error

Solution [continue]

1. The evidence is

$$p(x) = \Sigma_{i=1}^{c} p(x|w_i)P(w_i) = \begin{cases} 1 & 0 \le x \le \frac{cr}{c-1} \\ 1/c & i \le x \le i+1-\frac{cr}{c-1} \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

Because the $P(w_i)$ are constant, we have $P(w_i|x) \propto p(x|w_i)$ and thus

$$P(w_i|x) = \begin{cases} \frac{P(w_i)}{p(x)} = \frac{1/c}{p(x)} & 0 \le x \le \frac{cr}{c-1} \\ \left.\begin{matrix} 0 & \text{if } i \ne j \\ 1 & \text{if } i = j \end{matrix}\right\} & j \le x \le j+1-\frac{cr}{c-1} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

Ex.2: Nearest neighor error v.s. Bayes error

Solution [continue]

1. The conditional Bayesian probability of error at a point x is

$$P^*(e|x) = 1 - P(w_{\max}|x) \tag{13}$$
$$= \begin{cases} 1 - \frac{1/c}{p(x)} & \text{if } 0 \le x \le \frac{cr}{c-1} \\ 1 - 1 = 0 & \text{if } i \le x \le i + 1 - \frac{cr}{c-1} \end{cases} \tag{14}$$

and to calculate the full Bayes probability of error, we integrate as

$$P^* = \int P^*(e|x)p(x)dx \tag{15}$$
$$= \int_0^{cr/(c-1)} \left[1 - \frac{1/c}{p(x)}\right] p(x)dx \tag{16}$$
$$= \left(1 - \frac{1}{c}\right)\frac{cr}{c-1} = r. \tag{17}$$

Ex.2: Nearest neighbor error v.s. Bayes error

Solution [continue]

2. The nearest-neighbor error rate is

$$P = \int p(e|x)p(x)dx \tag{18}$$

$$= \int \Big[1 - \sum_{i=1}^{c} P^2(w_i|x)\Big]p(x)dx \tag{19}$$

$$= \int_0^{cr/(c-1)} \Big[1 - \frac{c(\frac{1}{c})^2}{p^2(x)}\Big]p(x)dx + \sum_{j=1}^{c} \int_j^{j+1-\frac{cr}{c-1}} [1-1]p(x)dx$$

$$= \int_0^{cr/(c-1)} \Big(1 - \frac{1/c}{p^2(x)}\Big)p(x)dx \tag{20}$$

$$= \int_0^{cr/(c-1)} \Big(1 - \frac{1}{c}\Big)dx = \Big(1 - \frac{1}{c}\Big)\frac{cr}{c-1} = r. \tag{21}$$

Thus we have demonstrated that $P^* = P = r$ in this nontrivial case.

Reference on Nearest Neighbor Error Bound

1. Cover, Thomas, and Peter Hart. Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 1967.
2. Duda, Richard O., Peter E. Hart, and David G. Stork. Pattern classification. 2012.