Non-parametric density estimation

Milos Hauskrecht
milos@pitt.edu
5329 Sennott Square

Nonparametric Density Estimation

- **Parametric distribution models** are:
  - restricted to specific functional forms, which may not always be suitable;
  - **Example:** modelling a multimodal distribution with a single, unimodal model.

- **Nonparametric approaches:**
  - Do not make any strong assumptions about the overall shape of the distribution being modelled.
Nonparametric Methods

**Histogram methods:**
partition the data space into distinct bins with widths $\Delta_i$ and count the number of observations, $n_i$, in each bin.

$$p_i = \frac{n_i}{N\Delta_i}$$

- Often, the same width is used for all bins, $\Delta_i = \Delta$.
- $\Delta$ acts as a smoothing parameter.
- Binning does not work well in a $d$-dimensional space.

---

Nonparametric Methods

- Binning does not work well in the in a $d$-dimensional space,
  - $M$ bins in each dimension will require $M^d$ bins!
- **Solution:**
  - Build the estimates of $p(x)$ by considering the data points in $D$ and how similar (or close) they are to $x$
  - **Example: Parzen window**
    - As if we build a bin dynamically for $x$ for which we need $p(x)$
Nonparametric Methods

- Assume observations drawn from a density \( p(x) \) and consider a small region \( R \) containing \( x \) such that
  \[
P = \int_{R} p(x) \, dx
\]

- The probability that \( K \) out of \( N \) observations lie inside \( R \) is \( \text{Bin}(K,N,P) \) and if \( N \) is large
  \[
  K \approx NP
  \]

If the volume of \( R \), \( V \), is sufficiently small, \( p(x) \) is approximately constant over \( R \) and

\[
P \approx p(x)V
\]

Thus

\[
p(x) = \frac{P}{V}
\]

Putting things together we get:

\[
p(x) = \frac{K}{NV}
\]

Nonparametric methods: kernel methods

**Solution 1:** Estimate the probability for \( x \) based on the fixed volume \( V \) built around \( x \)

\[
p(x) = \frac{K}{NV}
\]

- Fix \( V \), estimate \( K \) from the data

**Example:** Parzen window
Nonparametric methods: kernel methods

Kernel Density Estimation:

- **Parzen window**: Let \( R \) be a hypercube centred on \( x \) that defines the kernel function:

\[
k\left(\frac{x - x_n}{h}\right) = \begin{cases} 
1 & |(x_i - x_n)| / h \leq 1/2 \\
0 & \text{otherwise}
\end{cases}, \quad i = 1, \ldots, D
\]

- It follows that

\[
K = \sum_{n=1}^{N} k\left(\frac{x - x_n}{h}\right)
\]

- and hence

\[
p(x) = \frac{K}{NV} = \frac{1}{Nh^D} \sum_{n=1}^{N} k\left(\frac{x - x_n}{h}\right)
\]

Nonparametric Methods: smooth kernels

To avoid discontinuities in \( p(x) \) because of sharp boundaries we can use a smooth kernel, e.g. a Gaussian

- Any kernel such that

\[
h \cdot \text{acts as a smoother.}
\]
Nonparametric Methods: kNN estimation

**Solution 2:** Estimate the probability for \( x \) based on a fixed count \( K \) for a variable volume \( V \) built around \( x \)

**fix \( K \), estimate \( V \) from the data**

**Nearest Neighbour Density Estimation:**

Consider a hyper-sphere centred on \( x \) and let it grow to a volume, \( V^* \), that includes \( K \) of the given \( N \) data points. Then

\[
p(x) \simeq \frac{K}{NV^*}.
\]

---

Nonparametric vs Parametric Methods

**Nonparametric models:**

- More flexibility – no density model is needed
- But require storing the entire dataset
- And the computation is performed with all data examples.

**Parametric models:**

- Once fitted, only parameters need to be stored
- They are much more efficient in terms of computation
- But the model needs to be picked in advance