Non-parametric classification methods

Milos Hauskrecht
milos@cs.pitt.edu
5329 Sennott Square

Nonparametric vs Parametric Methods

Nonparametric models:
- More flexibility – no parametric 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
Non-parametric Classification methods

• Given a data set with \( N_k \) data points from class \( \mathcal{C}_k \) and \( \sum_k N_k = N \), we have

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

• and correspondingly

\[
p(x|\mathcal{C}_k) = \frac{K_k}{N_k V}.
\]

• Since \( p(\mathcal{C}_k) = N_k/N \), Bayes’ theorem gives

\[
p(\mathcal{C}_k|x) = \frac{p(x|\mathcal{C}_k)p(\mathcal{C}_k)}{p(x)} = \frac{K_k}{K}.
\]

K-Nearest-Neighbours for Classification
Nonparametric kernel-based classification

- **Kernel function:** $k(x, x')$
  - Models similarity between $x, x'$
  - **Example:** Gaussian kernel we used in the kernel density estimation
    \[
    k(x, x') = \frac{1}{(2\pi h^2)^{D/2}} \exp\left(-\frac{(x - x')^2}{2h^2}\right)
    \]
    \[
    p(x) = \frac{1}{N} \sum_{i=1}^{N} k(x, x_i)
    \]
- **Kernel for classification**
  \[
  p(y = C_k \mid x) = \frac{\sum_{x' : y' = C_k} k(x, x')}{\sum_{x'} k(x, x')}
  \]