K-Nearest-Neighbor Classifiers

Reference: The Elements of Statistical Learning, by T. Hastie, R. Tibshirani, J. Friedman, Springer
K-Nearest-Neighbor Classifiers Framework

• Classifiers:
  - memory-based
  - require no model to be fit

• Given a query point $x_0$
  $\rightarrow$ find the $k$ training points $x_{(r)}, r = 1, \ldots, k$ closest in distance to $x_0$
  $\rightarrow$ classify using majority vote among the $k$ neighbors
K-Nearest-Neighbor Classifiers Framework

• Successful in large number of classification problems: handwritten digits, satellite number image scenes and EKG patterns
• Each class has many possible prototypes
• Decision boundary is very irregular
K-Nearest-Neighbor Classifiers Framework

- The decision boundary of a 15-nearest-neighbor classifier applied to the three-class simulated data.
- Decision boundary is fairly smooth compared to the lower panel (1-nearest-neighbor classifier).
K-Nearest-Neighbor Classifiers

Framework

• Close relationship between nearest-neighbor and prototype methods:

in 1-nearest-neighbor classification, each training point is a prototype
K-Nearest-Neighbor Classifiers

Example

• STATLOG project: used part of a LANDSAT image as a benchmark for classification

• Four heat-map images, two in the visible spectrum and two in the infrared, for an area of agricultural land in Australia
K-Nearest-Neighor Classifiers Example

- Each pixel has a class label from the 7-element set $\mathcal{G} = \{red \ soil, \ cotton, \ vegetation \ stubble, \ mixture, \ gray \ soil, \ damp \ gray \ soil\}$
- Determined manually by research assistants surveying the area
- Lower middle panel
  - shows actual land usage
  - shaded by different colors to indicate the classes
K-Nearest-Neighbor Classifiers

Example

• Objective:
  classify the land usage at a pixel, based on the information in the four spectral bands

• Produced the predicted map shown in the bottom right panel:

  1. Extracted an 8-neighbor feature map – the pixel itself and its 8 immediate neighbors

  

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>N</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>X</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>
K-Nearest-Neighbor Classifiers

Example

2. Done separately in four spectral bands, giving \((1 + 8) \times 4 = 36\) input features

3. Five-nearest-neighbors classification was carried out in this 36-dimensional feature space

- Resulting test error rate \(~9.5\%\)
- Among all the methods used in the STATLOG project, k-nearest-neighbors performed best
  \(\rightarrow\) decision boundaries in \(\mathbb{R}^{36}\) are quite irregular