Nearest Neighbor Predictors

Stephen Vardeman

Analytics Iowa LLC

ISU Statistics and IMSE

Approximate conditional means and probabilities

- Optimal theoretical predictors often involve conditional (on the input vector) mean response or probabilities for the *K* possible values
- Approximations to these based on training data provide methods useable in practice
- "Nearest neighbor" approximations provide a full spectrum of ultimate predictor complexities/flexibilities (of the type needed in matching complexity to training set information content)

Motivation

- One might naively hope to use approximations like

$$\mathrm{E}[y \mid \mathbf{x}] \approx \frac{1}{\#\ \text{of}\ \mathbf{x}_i = \mathbf{x}} \sum_{i\ \text{s.t.}\ \mathbf{x}_i = \mathbf{x}} y_i \quad \text{or} \quad P[y = a \mid \mathbf{x}] \approx \frac{1}{\#\ \text{of}\ \mathbf{x}_i = \mathbf{x}} \sum_{i\ \text{s.t.}\ \mathbf{x}_i = \mathbf{x}} I[y_i = a]$$

but they will rarely work in practice, since typically counts of matching inputs in the training set are very small (if not 1)

- Roughly speaking, one might want to replace $\mathbf{x}_i = \mathbf{x}$ above with $\mathbf{x}_i \approx \mathbf{x}$

- One means of doing this is built on *k*-nearest neighborhoods

$$n_k(\boldsymbol{x}) = \text{the set of } k \text{ inputs } \boldsymbol{x}_i \text{ in the training set closest to } \boldsymbol{x} \text{ in } \Re^p$$

k-nn approximations

- Based on this reasoning, one is led to *k*-nn approximations

$$\mathrm{E}[y \mid \mathbf{x}] \approx \frac{1}{k} \sum_{i \text{ s.t. } \mathbf{x}_i \in n_k(\mathbf{x})} y_i \quad \text{or} \quad P[y = a \mid \mathbf{x}] \approx \frac{1}{k} \sum_{i \text{ s.t. } \mathbf{x}_i \in n_k(\mathbf{x})} I[y_i = a]$$

- These lead to the approximately optimal SEL predictor

$$\hat{f}(\mathbf{x}) \approx \frac{1}{k} \sum_{i \text{ s.t. } \mathbf{x}_i \in n_k(\mathbf{x})} y_i$$

and approximately optimal 0-1 loss classifier

$$\hat{f}(\mathbf{x}) \approx \underset{a}{\operatorname{argmax}} \ \#i \text{ s.t. } \mathbf{x}_i \in n_k(\mathbf{x}) \text{ and } y_i = a$$

k-nn SEL prediction

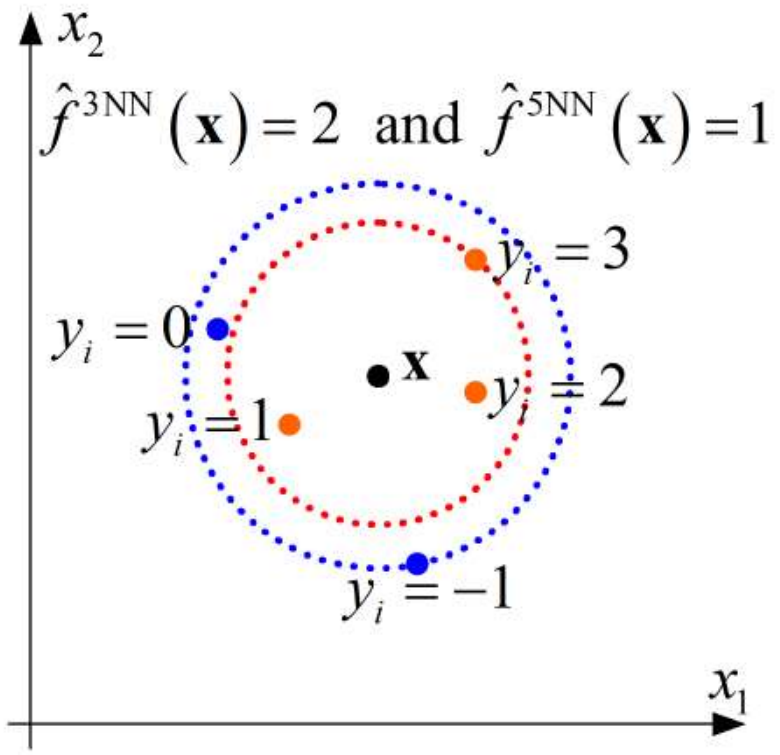
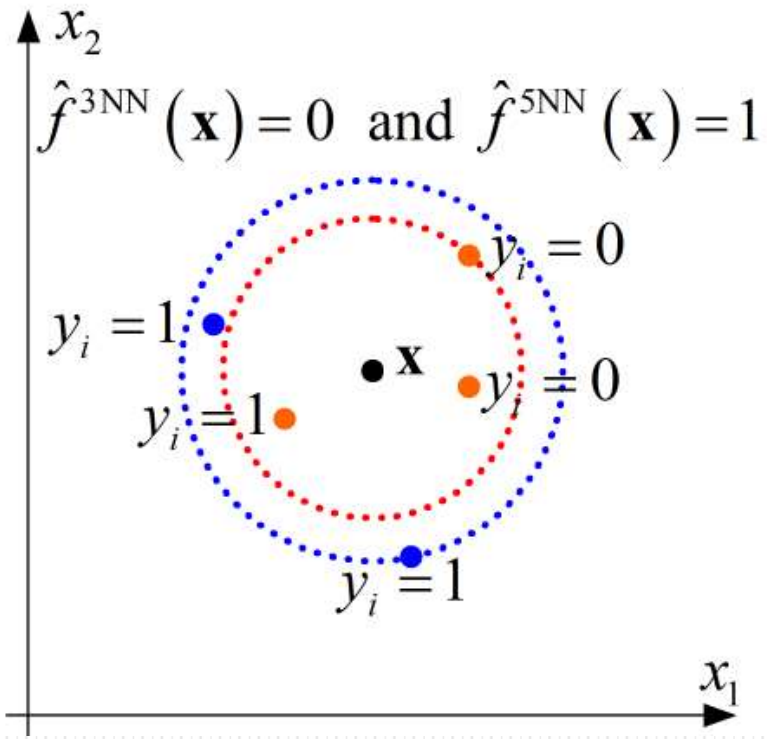


Figure: 5 Nearest Neighbors of $\mathbf{x}$ and $\hat{f}^{3NN}$ and $\hat{f}^{5NN}$

k-nn classification

Below is a cartoon illustrating 3- and 5-NN classification at **x** (in a 2-class problem with the $\{0, 1\}$ coding).



k-nn predictor complexity

- *k*-nn predictor complexity/flexibility decreases with increasing *k* (after all, *k*=1 simply predicts according to the single closest training case—a highly erratic proposition as one looks across the input space—while *k*=*N* produces constant predictors)

- The neighborhood size should thus be chosen to match complexity with real training set information content