Classifier Selection

James Loomis and Lei Hu

Classifier Selection Questions

- Build individual classifiers
- Evaluate competence of classifier for given x
  - Break ties for equal competences
- Selection Strategy
  - Take decision of classifier with highest competence
  - Fuse decision of most competent classifiers

Why Classifier Selection Works

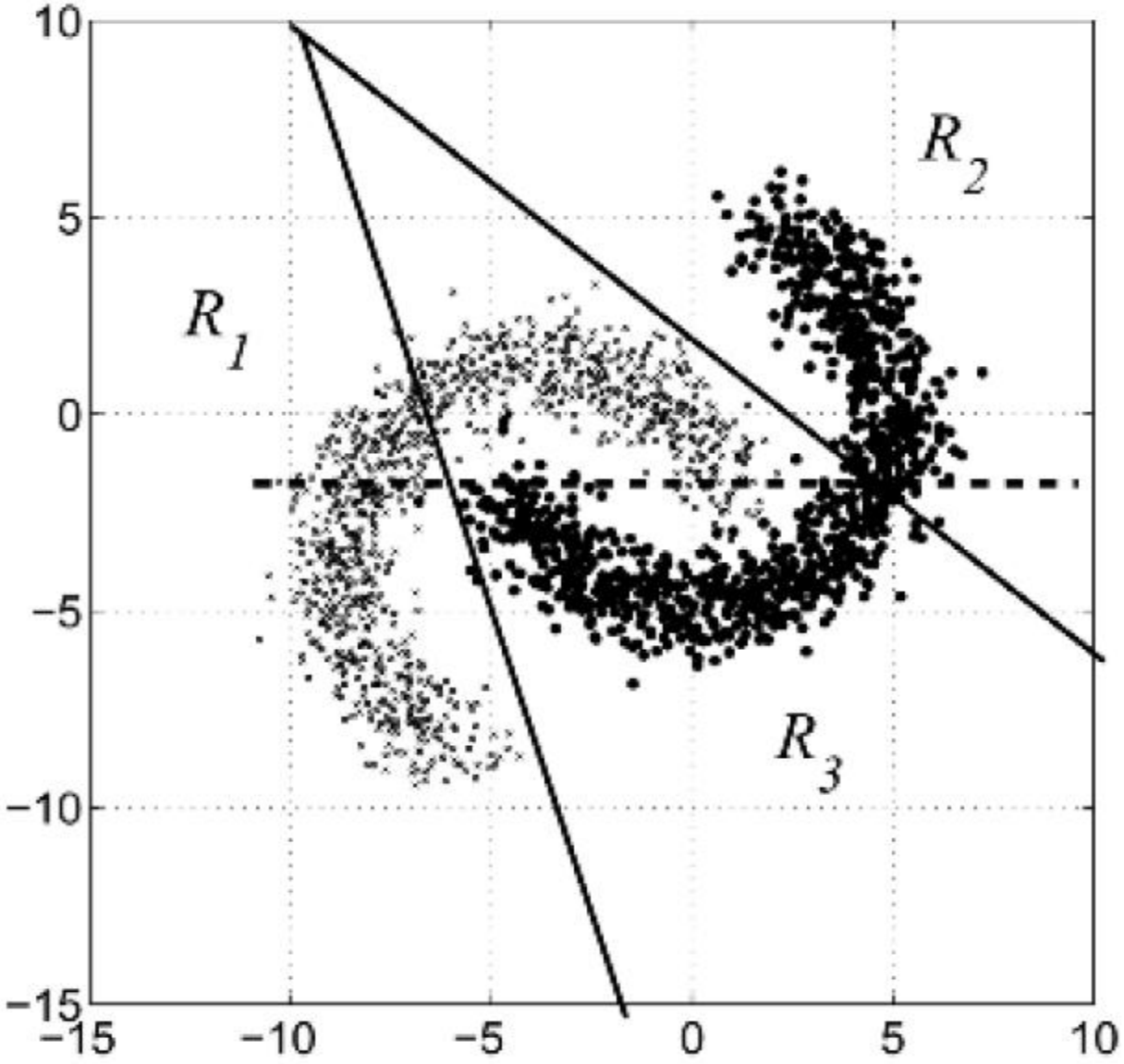

Fig. 6.1 *An example of partitioning the feature space with two classes into three s regions.*

Why Classifier Selection Works

$$P(correct) = \sum_{j=1}^{K} P(R_j) P(D_{i(j)} | R_j)$$

$$P(D_{i(j)} | R_j) \geq P(D_t | R_j), \forall\, t = 1, \ldots, L$$

$$P(correct) \gg \sum_{j=1}^{K} P(R_j) P(D^* | R_j) = P(D^*)$$

Decision-Independent Estimates

- Direct k-nn
  - Check accuracy on *K* nearest neighbors of **x**
- Distance-based k-nn

$$C(D_i|x) = \frac{\sum_{z_j \text{ in } N_x} P_i(l(z_j)|z_j)(1/d(x,z_j))}{\sum_{z_j \text{ in } N_x}(1/d(x,z_j))}$$

Decision-Dependent Estimates

- Direct k-nn

$$s_i \text{ in } \Omega$$

  - Choose *K* nearest neighbors of **x**, which are classified with same label.
  - Competence is proportion of those which are correct.

- Distance-based k-nn

$$C(D_i|x) = \frac{\sum_{z_j \text{ in } N_x} P_i(s_i|z_j)(1/d(x,z_j))}{\sum_{z_j \text{ in } N_x}(1/d(x,z_j))}$$

Tie Breaking Procedure

1. If all classifiers agree – return that label. Done.
2. Otherwise estimate local competence for each classifier, D_i.
3. If one with max. competence, allow it to label. If more than 1 tie, and there is a common consensus, return that label.
4. Have another competence contest. 2nd place winner chooses, if possible. Otherwise, random between remaining choices.
5. Finally, if we are here, just pick something random.

Preestimation of The Competence of Regions

- Precalculating the competence across regions of competence could reduce the computational amount.
- Clustering could ensure there are enough points in each region.

Clustering and Selection (training)

Clustering and selection (training)

1. Design the individual classifiers $D_1, \ldots, D_L$ using the labeled data set $\mathbf{Z}$. Pick the number of regions K.
2. Disregarding the class labels, cluster $\mathbf{Z}$ into K clusters, $C_1, \ldots, C_K$, using, e.g., the K-means clustering procedure [2]. Find the cluster centroids $\mathbf{v}_1, \ldots, \mathbf{v}_K$ as the arithmetic means of the points in the respective clusters.
3. For each cluster C_j, (defining region R_j), estimate the classification accuracy of $D_1, \ldots, D_L$. Pick the most competent classifier for R_j and denote it by $D_{i(j)}$.
4. Return $\mathbf{v}_1, \ldots, \mathbf{v}_K$ and $D_{i(1)}, \ldots, D_{i(K)}$.

Fig. 6.2 *Training of the clustering and selection method.*

Clustering and Selection (operation)

Clustering and selection (operation)

1. Given the input $\mathbf{x} \in \Re^n$, find the nearest cluster center from $\mathbf{v}_1, \ldots, \mathbf{v}_K$, say, $\mathbf{v}_j$.
2. Use $D_{i(j)}$ to label $\mathbf{x}$.

Fig. 6.3 *Operation of the clustering and selection method.*

Selective Clustering

- Selective Clustering provides a partition scheme for each classifier.
- Selective clustering could perform better than simple clustering and selection.

Selection or Fusion?

- Confidence interval and significantly best classifier is needed to avoid overtraining.
- Choosing selection if the best classifier is significantly better than the second best classifier, otherwise fusion

$$\Delta = 1.96\sqrt{\frac{P_1(1-P_1)}{N}} + 1.96\sqrt{\frac{P_2(1-P_2)}{N}}$$

Selection or Fusion?

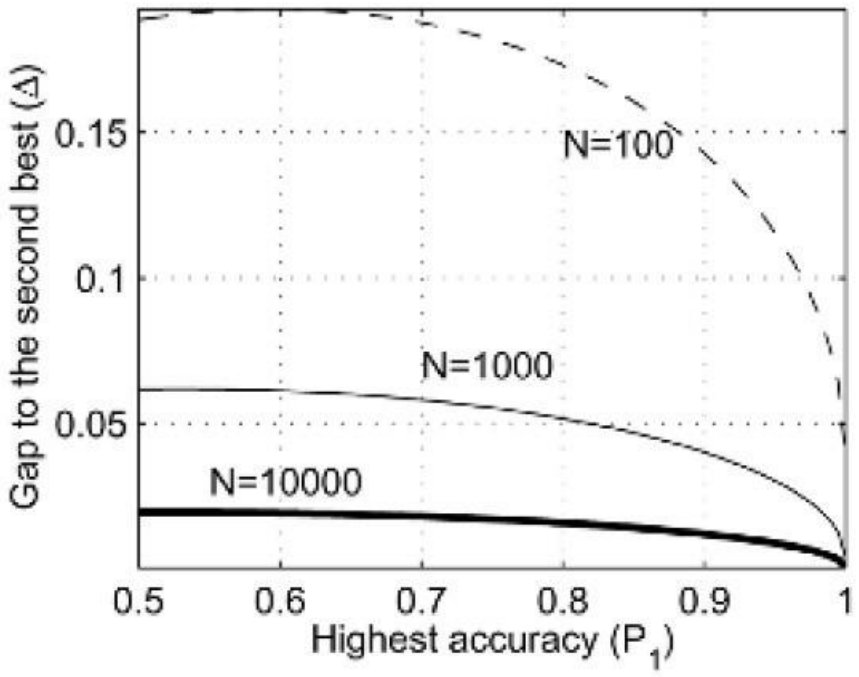

Fig. 6.5 *The difference Δ between the best and the second best classification accuracies in region R_j guaranteeing that the 95 percent CI of the two do not overlap.*

Base Classifiers and Mixture of Experts

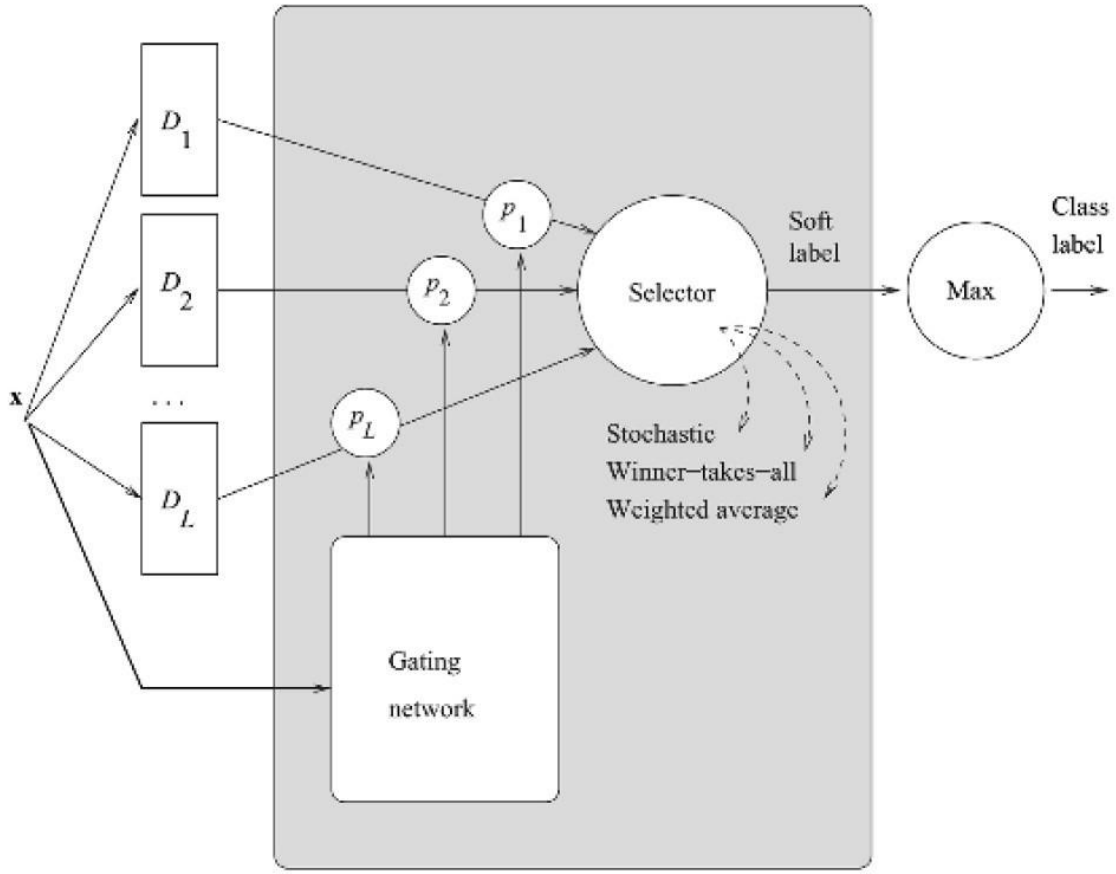

Fig. 6.6 *Mixture of experts.*

Mixture of Experts

- Stochastic selection
- Winner-takes-all
- Weights

$$\mu_j(x) = \sum_{i=1}^{L} p_i(x) d_{i,j}(x)$$

Training Procedure

- Standard error backpropagation
- Expectation maximization

References

- Kuncheva, L. (2004).

  Combining pattern classiers: methods and algorithms.

  Wiley-Interscience.

Questions?