Memory-Based Learning (MBL)/kNN

- Case-based, exemplar-based or instance-based learning
- A non-parametric method: no abstract model is created
- Training: just store the data (indexed!)
- Generalization: target function is estimated locally.
- The method is not aware of the global shape of the data, but does not make any assumptions about its shape
- Non-linear classifier
- "The cost for computation is decreasing roughly by a factor of two every year, whereas the price paid for incorrect assumptions is remaining the same." (Breiman and Friedman 1988)

MBL training and testing

- "We should never underestimate the power of simply remembering some or all of the examples and comparing test examples with our memory" (Ripley 1996) *Pattern Recognition and Neural Networks* ← a good book!
- Asymptotically, under various reasonable assumptions, risk converges to twice the Bayes risk as sample grows – independent of distance metric (Duda and Hart 1973; Ripley 1996)
  - Bayes risk = optimal error rate, assuming that probability distribution is known

MBL: similarity function

- Need distance function to calculate similarity
- In practice the metric is important
- Usually use Euclidean metric (alternatives Mahalanobis distance, L_1 distance = features different, etc.)
- Appropriate scaling of the variables first is important: normalize values, sometimes log-scale etc. Can put knowledge into metric.

MBL

- For 1NN decision surface is Voronoi diagram.
- Commonly distance-weight the classes of the neighbors in reaching a decision
  - e.g., information gain weighting used in TiMBL (Tilburg Memory-based Learner http://ilk.kub.nl/software.html)
- Asymptotically this doesn't help, but can allow use of larger k for fixed size training set
- $k > 1$ increases resistance to noise (decreases variance)

Feature selection for MBL

- Excluding features which have little or no relevance is very important
- Wrapper methods generate candidate features by testing performance of the classifier with those features
- Filter methods decide features independently of algorithm that will use them
- Can use decision tree techniques to select features (Cardie 1993)

MBL

- Naive storage and search for neighbors can be expensive
- But there are efficient algorithms for indexing and hence finding neighbors (k-d trees)
- The stored data set size can be reduced by data editing so as to only store examples near decision surfaces.