TTT4275 Summary on classification

Spring 2018

Faglærer: Magne Hallstein Johnsen,

Institutt for elektronikk og telekommunikasjon, NTNU

NTNU

Regions and borders and optimal classifier

- Borders/regions depend on chosen features
- The **true** class regions in the input room are unknown and overlap (nonseparable)
- The **decision** borders and regions are given by the specific classifier and do not overlap.
- The decision regions/borders are different from the **true** class regions/borders.
- Theoretically optimal recognizer is given by the Bayes decision rule (BDR).
  - $x \in \omega_k \Leftrightarrow P(\omega_k/x) = \max_i P(\omega_i/x)$ or :
  - $x \in \omega_k \Leftrightarrow p(x/\omega_k)P(\omega_k) = \max_i p(x/\omega_i)P(\omega_i)$
- $P(\omega_i), P(\omega_i/x), p(x/\omega_i) \quad i = 1, C$ are **never** known

Practical classifiers

- "Plug in BDR/MAP" classifier use a parametric model/approximation to the true/unknown class densities
  - GMM, i.e. $p(x/\omega_i) = \sum_{k=1}^{K} c_{ik} N(\mu_{ik}, \Sigma_{ik})$ is state-of-the-art
- Discriminant classifiers use a chosen function $g_i(x)$ for each class $i = 1, \ldots, C$
  - Nonlinear choices include neural networks (state-of-the-art) and machine learning
  - This course focus on the linear choice $g(x) = Wx$ (matrix form)
  - The linear choice is usually expanded to use binary outputs, i.e. $g(x) = sigmoide(Wx)$
- Reference based classifiers use training set or derived versions as (labeled) references
  - Classification is based on KNN/NN, i.e minimum distance(s) from input to references

Designing/training practical classifiers

- Based on labeled training set $X = \{X_1, X_2, \ldots, X_C\}$
- Plug-in-MAP is trained one class at a time, i.e $p(x/\omega_i) = p(x/\lambda_i)$ is trained by X_i
  - ML training: $max_{\lambda_i} p(X_i/\lambda_i)$ (using the EM-algorithm when GMM)
  - MAP training uses priors : $max_{\lambda_i} p(\lambda_i/X_i) = max_{\lambda_i} p(X_i/\lambda_i)p(\lambda_i)$
- The matrix W in linear classifiers is trained by gradient techniques using all X simultaneously.
  - Perceptron criteria if separable problem (seldom)
  - In practice by $MSSE = \sum_k (g_k - t_k)^T (g_k - t_k)$
  - Here t_k is the binary label vector for training vector $x_k \in X$
- References in the KNN/NN classifiers is usually found by clustering
  - X_i training vectors are clustered into a small sets of references for class ω_i using K-Means.
  - If X is small all training vectors can be used as references

Evaluation of classifiers

- Given a chosen classifier structure and two different **training** sets :
  - will result in two different classifiers, each with a true (but unknown) test error rate
  - only an infinitely large training set will give the structure-optimized test error rate
- For any trained classifier and two different **test** sets :
  - will result in two different (empirical) error rates, and both different from the true error rate
  - thus only an infinitely large test set will give the true error rate.
- For any trained classifier and a finite test set .
  - The true error rate can be estimated as a 95% confidence interval centered on the empirical test error rate
  - The interval size is dependent on the empirical error rate and the size of the test set.
- The confusion matrix shows the true versus classified labels for the test set.