

This document is also available in PDF¹ format.

Purpose: To introduce non-parametric pdfs and related concepts.

Material: Section 4.2 in notes. Also Devijver&Kittler App. A.

General: Quite often we do not know the appropriate pdf shape needed in a specific situation. The techniques covered in this lecture allow for direct pdf estimation when a set of features for that pdf is available. These techniques are very useful when little is known about the features under consideration. The usual parametric versus non-parametric play-off is present here. Less assumptions are necessary here, but that takes place at the cost of increased computation and the loss of smoothing characteristics which parametric methods offer. As we shall see in later lectures, neural nets, which are very flexible parametric models, are often also useful in situations where non-parametric pdf estimations are used.

Topics:

- Equation 4.31 in the notes provides the basis for the following techniques. Remember that this equation is applied per class.
- *Histograms:* Divide the feature space into rectangular bins of equal fixed volume. After counting the number of feature vectors in each bin, the pdf can be approximated by a piece-wise constant, pre-calculated function. Seldom used for PR; try to imagine the data structure necessary to save a pdf approximation for 20-dimensional feature vectors, where the range of each feature in the vector is represented on 10 intervals!
- *Parzen:* Estimate the pdf on the fly. Choose a volume around the feature vector in question and count the data in it. Several volume shapes are available, typically hypercube (fast but spiky) or hypersphere. Using Gaussian kernels gives a smooth function, but is very computationally intensive (related to the RBF). The volume of a hypersphere is given by

$$V(D, r) = \frac{\pi^{D/2} r^D}{\Gamma(D/2 + 1)}$$

where D is the dimension and r is the radius.

- *K-nearest neighbours:* Find the K nearest vectors to the point where the pdf must be estimated. Then calculate the volume necessary to enclose them. Results in a smoother estimate. (A variant on this is the popular K -NN rule used for classification. This is however applied on all the data simultaneously and gives posterior class probability estimates, not a class specific pdf estimation).
- Another PR subject which is bubbling under the surface here, and which closely ties in with classification, is *distance measures*. Watch out for names like city-block, Chebychev, Euclidean, Mahalanobis etc. See Devijver & Kittler for details.
- Somewhere between non-parametric and parametric pdfs, there are mixture Gaussian densities (GMMs), radial basis function networks (RBFs) and multi-layer perceptrons (MLPs). These will be considered in later lectures.

Project: (SELF-STUDY)

- Repeat the previous (parametric) classification task, this time using any non-parametric pdf estimate of your choice. Compare results.

¹http://www.dsp.sun.ac.za/pr813/lectures/4_nonparam/4_nonparam.pdf