# Indicative Exercises – Problems for the Pattern Recognition direction of the course Singal/Image Processing and Pattern Recognition

1. (a) Define the classification and the regression problems.
   (b) Consider a set of points that lie around a curve. Propose a parametric model to represent the data. Use this model for prediction.
   (c) Consider a set of points that form compact clusters around two given points. Propose a parametric model that models the data. Use this model for classification.

2. Consider a two-class problem (in the 1-dim. space) where each class is modeled by two normal pdfs, $p(x|\omega_1)$ =N($m_1$,$s_1^2$) and $p(x|\omega_2)$=N($m_2$,$s_2^2$) and having a priori probabilities $P(\omega_1)$ and $P(\omega_2)$, respectively. Determine the decision regions for each class.
   Example: $s_1$=$s_2$ and $P(\omega_1)$=2/3, $P(\omega_2)$=1/3.

3. Consider a 1-dim., two-class classification problem where $p(x|\omega_1)$ =N(*2*,*4*) and $p(x|\omega_2)$=N(-*2*,*4*). Classify x=1 to a class when (a) $P(\omega_1)$=$P(\omega_2)$ and (b) $P(\omega_1)$=1/3, $P(\omega_2)$=2/3.

4. Consider a 1-dim. two-class classification problem where the two classes are modeled by uniform distributions $p(x|\omega_1)$ =1/6, if $x\in(3,9)$ and 0 otherwise and $p(x|\omega_1)$ =1/3, if $x\in(5,8)$ and 0 otherwise. For each of the scenarios (i) $P(\omega_1)$= $P(\omega_2)$ and (ii) $P(\omega_1)$=5/6 and $P(\omega_2)$=1/6, perform the following steps:
   (a) Plot in the same graph $P(\omega_1)p(x|\omega_1)$ versus $x$ and $P(\omega_2)p(x|\omega_2)$ versus $x$.
   (b) Identify the decision regions for each class.
   (c) Compute the probability of classification error.

5. Under what conditions the Euclidean classifier is optimum wrt the probability of classification error criterion?

6. Consider the Erlang distribution $p(x)=\vartheta^2 x \exp(-\vartheta x)u(x)$, whose mean equals to $1/\vartheta$.

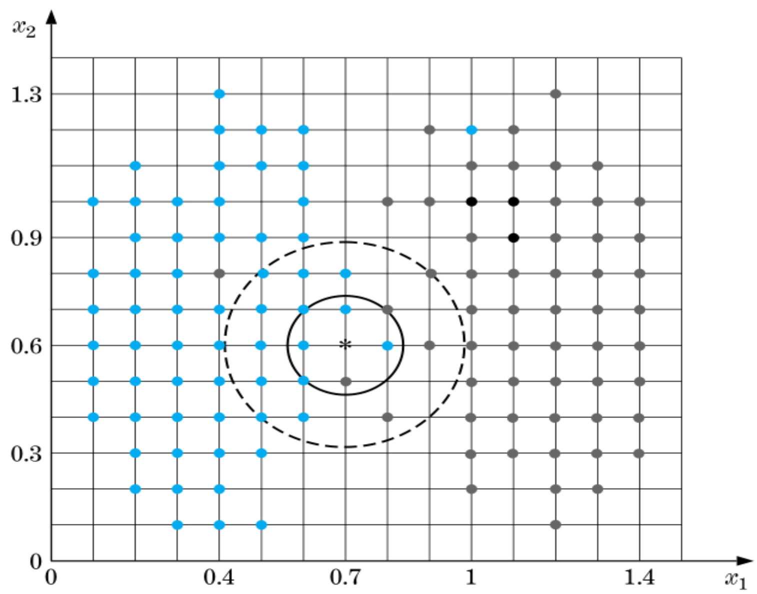
   (a) Given a set of $N$ measurements $x_1,\ldots,x_N$, for the random variable $x$ that follows the Erlang distribution, prove that the ML estimate of $\vartheta$ is
   $$\vartheta_{ML}= 2N / \sum_{i=1}^{N} x_i$$
   (b) For $N$=5 and $x_1$=2, $x_2$=2.2, $x_3$=2.7, $x_4$=2.4, $x_5$=2.6, estimate the mean of the random variable $x$.

7. Suppose that in a classification problem it is known that the (**known**) data from one class form two compact groups around two (**unknown**) points in space. What model

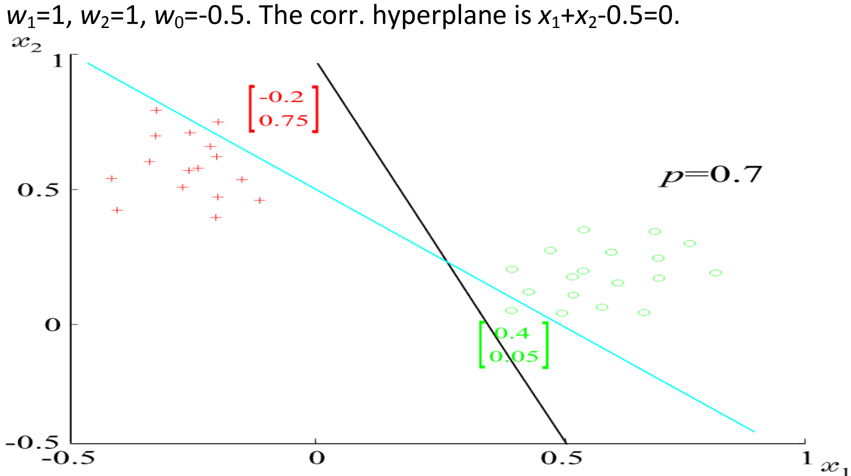
would you choose to represent the pdf of this class? How would you identify its parameters?

8. When the naïve Bayes classifier is equivalent to the Bayes classifier? What is the main feature of this algorithm that makes it popular?

9. Given a set of points from two classes, classify the given data vector to a class using the Bayes rule, where the pdfs of the classes are estimated via the k-NN density estimation.



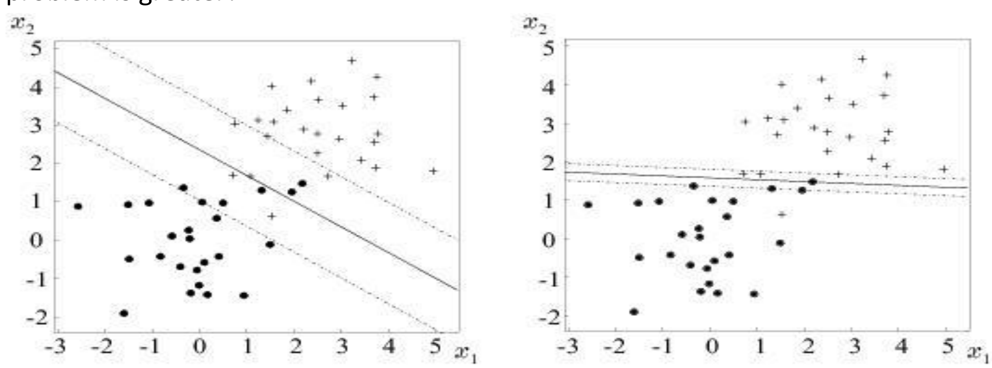

10. Perform a single iteration of the perceptron algorithm (Remember to augment the involved data vectors by an additional coordinate equal to 1).
**Example:** At some stage *t* the perceptron algorithm results in which corresponds to $w_1$=1, $w_2$=1, $w_0$=-0.5. The corr. hyperplane is $x_1+x_2-0.5=0$.




$$\underline{w}(t+1)=\begin{bmatrix}1\\1\\-0.5\end{bmatrix}-0.7(-1)\begin{bmatrix}0.4\\0.05\\1\end{bmatrix}-0.7(+1)\begin{bmatrix}-0.2\\0.75\\1\end{bmatrix}=\begin{bmatrix}1.42\\0.51\\-0.5\end{bmatrix}$$

11. Given a set of points from two classes, determine the LS classifier that is based on the sum of error squares criterion $\hat{w} = (X^T X)^{-1} X^T y$
(Remember to augment the involved data vectors by an additional coordinate equal to 1).

12. Determine the linear SVM classifier for a symmetric set of data from two linearly separable classes. In which of the following two cases the parameter C of the SVM problem is greater?



(a) (b)

13. Consider the lines (e1) $x_1$=0, (e2) $x_2$=0 and (e3) $x_1$+$x_2$=2, which all leave the point (0.5,0.5) from their positive side. Assume that all the points that lie in the positive side of all lines belong to class A, while all the remaining belong to class B. Construct a multilayer perceptron that implements the above classification.

14. What is the reason for replacing the hard limiter function in the nodes of the multilayer perceptrons (MLPs) with smoother ones?

15. Does the Back propagation algorithm always find the best possible solution for the classification problem at hand? Explain.

16. Propose a method for avoiding overtraining.

17. “A MLP classifies perfectly all points of the training set. This assures that the network has high generalization ability.” Do you agree with this statement? Explain.

18. What is the main strategy that is followed by several nonlinear classifiers?

19. “Every mapping of the data to a higher dimensional space surely makes the data more linearly separable.” Do you agree with this statement? Explain.

20. Let $\underline{x} = [x_1, x_2]^T \in R^2 and\ \underline{x} \rightarrow \underline{y} = \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix} \in R^3$ . Prove that $\underline{y}_i^T \underline{y}_j = (\underline{x}_i^T \underline{x}_j)^2$