Recognition January 10, 2013

Name:

1. Objects from two classes $k \in \{1, 2\}$ are classified given a single observation, a non-negative real number x. Conditional probability densities $p(x|k)$ are: $p(x|1) = 1, x \in (0, 1)$ and 0 elsewhere, $p(x|2) = \exp(-x)$. Find the decision d optimal for the
   (a) (2 points) the minmax formulation of the recognition task.
   (b) (2 points) the Neyman-Pearson formulation of the recognition task. The false negative rate (probability of misclassifying a dangerous object) is required to be below 0.1 (10%). Class 1 is considered dangerous.
   (c) (2 points) the Bayesian recognition task, for equal a priori probabilities and the 0-1 loss matrix (i.e. for the minimisation of the probability of incorrect decision)
   (d) (2 points) the Bayesian recognition task, for equal a priori probabilities and the loss matrix W of the form (row: true class, column: decision)

   $$W = \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}.$$

   (e) (2 points) the Bayesian recognition task with a reject option ("don't know" decision), for equal a priori probabilities and the loss matrix W of the form (row: true class, column: decision):

   $$W = \begin{bmatrix} 0 & 1 & \varepsilon \\ 1 & 0 & \varepsilon \end{bmatrix}.$$

   Consider $\varepsilon = 0.1$ and $\varepsilon = 0.2$. Bonus (1 point): consider ε a free parameter and express the decision as its function.

2. (5 points) Describe the AdaBoost learning algorithm. Include the formula for setting α_i and justify it. Discuss the time complexity of the algorithm. Compare its properties with SVM, neareast neighbour and neural network learning and classification methods.

3. The K-means algorithm.
   (a) (1 point) Describe the basic algorithm.
   (b) (1 point) Your implementation of the K-means algorithm is run for $K = 5$ and $K = 6$ and the value of the cost function, i.e. the sum of squared difference from the prototypes is lower for $K = 6$ than for $K = 5$. Is this possible with a correct implementation ?
   (c) (1 point) Describe a generalization to L_1 minimization, i.e. to the minimization of the sum of absolute differences.
   (d) (1 point) Describe how to use k-means for string clustering.
   (e) (1 point) Discuss the time complexity of various implementations of the standard algorithm and the generalizations.