Recognition January 23, 2018

Name:

1. (5 points) Neyman-Pearson.
   (a) (2 points) Formulate the Neyman-Pearson problem.
   (b) (3 points) Find the optimal Neyman-Pearson strategy for the following problem. Feature vectors are real numbers $x \in [0, 1]$. Objects are from two classes $k \in \{1, 2\}$. The conditionals $p(x|k)$ are:

$$p(x|1) = 2x\,, \tag{1}$$

$$p(x|2) = 3x^2\,. \tag{2}$$

   Class 1 is considered dangerous. The limit ϵ_1 for maximum overlooked danger is 0.1 (10%).

2. (5 points) Describe the Adaboost learning algorithm. Discuss the time complexity of the algorithm. Compare its properties with SVM, Logistic regression, the neareast neighbour and neural network learning and classification methods.

3. (6 points) Maximum likelihood estimation.
   (a) (4 points) The reliability of light bulbs is defined by the probability density function $p(t) = 1/\theta\, e^{-t/\theta}, t \in [0, \infty]$, where $p(t)\delta$ is the probability that a bulb fails in a short interval δ centered at time t. What is the maximum likelihood estimate of the reliability parameter θ if in an experiment with six bulbs the following lifetimes have been observed:
   $t_0 = 0.0, t_1 = 1.0, t_2 = 2.3, t_3 = 2.7, t_4 = 6.0, t_5 = 12.0$?
   (b) (2 points + 1 bonus point) If the test decribed above finished at time $t_{end} = 4.0$ two light bulbs would still be on, i.e. their time of failure would be unknown. What would an ML estimate for θ be in this case?
   Hints: There are two possible ways to solve the problem. Either solve the ML estimation directly, using probability of the light bulbs still functioning at time t_{end} for the uknown failure times. Alternatively, use the EM algorithm which will "substitute" for the unknown failure times the expectation of the failure time after t_{end}.

4. (4 points) Decision tree for deciding whether or not to go to the cinema. The attributes of training data are: (1) movie length (short/long), (2) country of origin (American/other), (3) the weather (nice/rainy).
   Find the first splitting atribute by maximization of the information gain (IG). Consider the following training data (note that attributes are all binary, target decision is also binary.)

Training sample index	Short	American	Nice weather	Decision: Go?
1	1	1	0	1
2	1	0	0	1
3	1	1	1	0
4	0	1	0	1
5	1	0	1	0
6	0	0	0	0
7	1	1	1	0
8	1	0	1	1

   Hint: $\log_2 3 \approx 1.6$