Recognition January 15, 2019

Name:

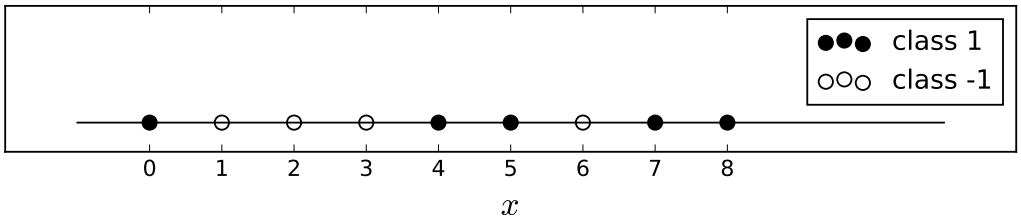
1. (8 points) Maximum Likelihood Estimation.

   (a) (3 points) The reliability of light bulbs is defined by the probability density function $p(t) = \frac{1}{\theta} e^{-t/\theta}, t \in [0, \infty]$, 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) (3 points) 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? Solve the ML estimation directly using probability of the light bulbs still functioning at time t_{end} for the uknown failure times (2 points). Also do this using EM algorithm which will "substitute" for the unknown failure times the expectation of the failure time after t_{end} (1 point).

   (c) (2 points) The same problem, but the device for logging the failure times is not available. The only information we have is that at time t_{end}, G bulbs are still good and F bulbs have already failed. Find θ.

2. (4 points) Adaboost learning algorithm. Consider the following 1-D data:



   and the following set of weak classifiers: $h(x) = \text{sign}(ax + b)$ $(a, b \in \mathbb{R})$. Use this example to explain how Adaboost works (make one full iteration, ending with first data re-weighting.)

3. (4 points) k-nearest neighbors. Describe the algorithm and list its pros and cons.

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: $\log_2 3 \approx 1.6$