RPZ Homework

November 15, 2018

Bayesian Decision Theory

A student prepares for the exam. There are K tickets in total, one for each lecture. Because the lectures are sequential, he prepares sequentially. He learns the first ticket with probability q. If he already learned k tickets, he learns the next one with probability q or otherwise stops preparing.

At the exam a ticket is drawn randomly. Assume the student answered well the ticket with number x. The task is to recognize whether he has prepared at least half of the tickets (assume K is even). Model the problem as a Bayesian decision.

1. In this problem, what is the hidden state, observation, decision?
2. What is the probability that he learned at least half of the tickets?
3. Derive the optimal Bayesian decision strategy.

Logistic Regression with Label Errors

Suppose that the class label k given an observation x follows the logistic model with conditional distribution $q(k|x;w)$, where $k \in \{-1, 1\}$ and w are model parameters. Suppose you have training pairs (t_i, x_i) where t_i might have been incorrectly labelled, which happens with probability ε. That is $t_i = -k_i$ with probability ε and $t_i = k_i$ with probability $1 - \varepsilon$, where k_i is the true label which is not available.

1. Formulate maximum likelihood learning of this model.
2. How the maximum likelihood is related to minimizing the cross-entropy $\sum_i \sum_k p_i(k) \log q(k \mid x_i; w)$ with $p_i(k) = 1 - \varepsilon$ for $k = t_i$ and ε otherwise? Hint: show that cross-entropy is an upper bound on the negative log likelihood.
3. Show that minimizing the cross-entropy is equivalent to minimizing the KL divergence from p to q.