Artificial Intelligence II

In this question we deal with a general two-class supervised learning problem. Instances are denoted by \( x \in X \), the two classes by \( c_1 \) and \( c_2 \), and \( h : X \to \{c_1, c_2\} \) denotes a hypothesis. Labelled examples appear independently at random according to the distribution \( P \) on \( X \times \{c_1, c_2\} \). The loss function \( L(c_i, c_j) \) denotes the loss incurred by a classifier predicting \( c_i \) when the correct prediction is \( c_j \).

(a) Show that if our choice of hypothesis \( h \) is completely unrestricted and \( L \) is the 0–1 loss function then the Bayes optimal classifier minimising

\[
E[L(h(x), c)]
\]

where the expected value is taken according to the distribution \( P \) is given by

\[
h(x) = \begin{cases} c_1 & \text{if } Pr(c_1|x) > \frac{1}{2} \\ c_2 & \text{otherwise}. \end{cases}
\]

[10 marks]

(b) We now define a procedure for the generation of training sequences, denoted by \( s \). Let \( \mathcal{H} \) be a set of possible hypotheses, let \( p(h) \) be a prior on \( \mathcal{H} \), let \( p(x) \) be a distribution on \( X \) and let \( Pr(c|x, h) \) be a likelihood, denoting the probability of obtaining classification \( c \) given instance \( x \) and hypothesis \( h \in \mathcal{H} \). A training set \( s \) is generated as follows. We obtain a single \( h \in \mathcal{H} \) randomly according to \( p(h) \). We then obtain \( m \) instances \((x_1, \ldots, x_m)\) independently at random according to \( p(x) \). Finally, these are labelled according to the likelihood such that

\[
p(s|h) = \prod_{i=1}^{m} Pr(c_i|x_i, h)p(x_i).
\]

We now wish to construct a hypothesis \( h' \), not necessarily in \( \mathcal{H} \), for the purposes of classifying future examples. The usual approach in a Bayesian context would be to construct the hypothesis

\[
h'(x) = \begin{cases} c_1 & \text{if } Pr(c_1|x, s) > \frac{1}{2} \\ c_2 & \text{otherwise}. \end{cases}
\]

By modifying your answer to part (a) or otherwise, show that this remains an optimal procedure in the case of 0–1 loss.

[10 marks]