Homework 6 Solutions

Problem 1.

(a) Let H_k represent the hypothesis that the bag contains k black marbles and one red one. The MAP method chooses the hypothesis H_k that maximizes the posterior probability of the data:

$$\Pr\{H_k|D\} = \frac{\Pr\{D|H_k\} \cdot \Pr\{H_k\}}{\Pr\{D\}}$$

Note that for all k, the denominator ($\Pr\{D\}$) is the same, so for our purposes it can be considered a constant. Thus, we need only find a k that maximizes $\Pr\{D|H_k\} \Pr\{H_k\}$.

From the definition of the problem, we can derive the following probabilities:

$$\Pr\{R|H_k\} = \frac{1}{k+1}$$

$$\Pr\{BBBR|H_k\} = \frac{k}{k+1} \cdot \frac{k+1}{k+2} \cdot \frac{k+2}{k+3} \cdot \frac{1}{k+4} = \frac{k}{(k+3)(k+4)}$$

$$\Pr\{BR|H_k\} = \frac{k}{k+1} \cdot \frac{1}{k+2}$$

$$\text{Thus, } \Pr\{D|H_k\} = \Pr\{R|H_k\} \cdot \Pr\{BBBR|H_k\} \cdot \Pr\{BR|H_k\},$$

$$= \frac{k^2}{(k+1)^2(k+2)(k+3)(k+4)}$$

$$\text{So } \Pr\{D|H_k\} \Pr\{H_k\} = \frac{k^2}{(k+1)^2(k+2)(k+3)(k+4)2^k}$$

This value is maximized when $k = 1$. To get full credit for this problem, you needed to prove that $k = 1$ maximizes this function for all $k \geq 1$, not just show that the function decreases for the first few values of k. An observation that the 2^k in the denominator will quickly dominate the expression was generally sufficient, but many other proofs were possible. One solution (due to Peter McCorquodale, in Fall '93) lists the values of the posterior probability for k between 1 and 3, and then notes that for $k \geq 3, (1 + \frac{1}{k})^2 \leq 2$. Therefore,

$$\Pr\{D|H_k\} \Pr\{H_k\} = \frac{1}{(1+\frac{1}{k})^2(k+2)(k+3)(k+4)2^k}$$

$$\geq \frac{1}{2(k+2)(k+3)(k+4)2^k}$$

$$> \frac{1}{(k+3)(k+4)(k+5)2^{k+1}} = \Pr\{D|H_{k+1}\} \Pr\{H_{k+1}\}.$$

(b) The MDL principle favors the "least complex" hypothesis, where complexity is measured in terms of the number of bits needed to describe both the hypothesis and the data. The description length can be expressed as

$$-\log \Pr\{H_k|D\} = -\log \frac{\Pr\{D|H_k\} \cdot \Pr\{H_k\}}{\Pr\{D\}}$$

$$= -\log \Pr\{D|H_k\} - \log \Pr\{H_k\} + \log \Pr\{D\}$$

Once again, since $\Pr\{D\}$ is the same for all k, we can ignore this term.

Since $\Pr\{H_k\} = 2^{-k}$, the description length of H_k is just k. So the optimal hypothesis in this model is the k that minimizes

$$k - \log \frac{k^2}{(k+1)^2(k+2)(k+3)(k+4)2^k}$$

$$= k + 2\log(k+1) + \log(k+2) + \log(k+3) + \log(k+4) - 2\log k.$$

Since this is just the log of the function we computed in part **a**, we know it is minimized when $\Pr\{D|H_k\} \cdot \Pr\{H_k\}$ is maximized; that is, when $k = 1$.