Massachusetts Institute of Technology

Department of Electrical Engineering & Computer Science

6.041/6.431: Probabilistic Systems Analysis

(Fall 2008)

Tutorial 11 - Solutions
November 20/21, 2008

1. According to the MAP rule, we maximize the posterior PDF over $\theta \in [0,1]$. We let X be the number of heads observed and proceed:

$$f_{\Theta|X}(\theta \mid k) = \frac{f_\Theta(\theta) p_{X|\Theta}(k \mid \theta)}{\int f_\Theta(\theta') p_{X|\Theta}(k \mid \theta')\, d\theta'},$$

where X is the number of heads observed. Since the denominator is a positive constant, we only need to maximize

$$f_\Theta(\theta) p_{X|\Theta}(k \mid \theta) = \binom{n}{k} \left(2 - 4\left|\frac{1}{2} - \theta\right|\right) \theta^k (1-\theta)^{n-k}.$$

The function to be minimized is differentiable except at $\theta = 1/2$. This leads to three different possibilities: (a) the maximum is attained at $\theta = 1/2$; (b) the maximum is attained at some $\theta < 1/2$, at which the derivative is equal to zero; (c) the maximum is attained at some $\theta > 1/2$, at which the derivative is equal to zero.

Let us consider the second possibility. For $\theta < 1/2$, we have $f_\Theta(\theta) = 4\theta$. The function to be maximized, ignoring the constant term $4\binom{n}{k}$, is

$$\theta^{k+1}(1-\theta)^{n-k}.$$

By setting the derivative to zero, we find $\hat{\theta} = (k+1)/(n+1)$, provided that $(k+1)/(n+1) < 1/2$. Let us now consider the third possibility. For $\theta > 1/2$, we have $f_\Theta(\theta) = 4(1-\theta)$. The function to be maximized, ignoring the constant term $4\binom{n}{k}$, is

$$\theta^k (1-\theta)^{n-k+1}.$$

By setting the derivative to zero, we find $\hat{\theta}_{\text{MAP}} = k/(n+1)$, provided that $k/(n+1) > 1/2$. If neither condition $(k+1)/(n+1) < 1/2$ and $k/(n+1) > 1/2$ holds, we must have the first possibility, with the maximum attained at $\hat{\theta}_{\text{MAP}} = 1/2$.

2. (a) Let K be the number of heads observed before the first tail, and let $p_{K|H_i}(k)$ be the PMF of K when hypothesis H_i is true. Note that event $K = k$ corresponds to a sequence of k heads followed by a tail, so that

$$p_{K|H_i}(k) = (1-q_i)q_i^k, \qquad k = 0, 1, \ldots, \quad i = 1, 2.$$

Using Bayes' rule, we obtain

$$\begin{aligned} \mathbf{P}(H_1 \mid K = k) &= \frac{p_{K|H_1}(k)\mathbf{P}(H_1)}{p_K(k)} \\ &= \frac{\frac{1}{2}(1-q_1)q_1^k}{\frac{1}{2}(1-q_1)q_1^k + \frac{1}{2}(1-q_0)q_0^k} \\ &= \frac{(1-q_1)q_1^k}{(1-q_1)q_1^k + (1-q_0)q_0^k}. \end{aligned}$$

(b) An error occurs in two cases: if H_0 is true and $K \geq k^*$, or if H_1 is true and $K < k^*$. So, the probability of error, denoted by p_e, is

$$\begin{aligned}
p_e &= \mathbf{P}(K \geq k^* \mid H_0)\mathbf{P}(H_0) + \mathbf{P}(K < k^* \mid H_1)\mathbf{P}(H_1) \\
&= \sum_{k=k^*}^{\infty} p_{K|H_0}(k)P(H_0) + \sum_{k=0}^{k^*-1} p_{K|H_1}(k)\mathbf{P}(H_1) \\
&= \mathbf{P}(H_0)\sum_{k=k^*}^{\infty}(1-q_0)q_0^k + P(H_1)\sum_{k=0}^{k^*-1}(1-q_1)q_1^k \\
&= \mathbf{P}(H_0)(1-q_0)\frac{q_0^{k^*}}{1-q_0} + \mathbf{P}(H_1)(1-q_1)\frac{1-q_1^{k^*}}{1-q_1} \\
&= \mathbf{P}(H_1) + \mathbf{P}(H_0)q_0^{k^*} - \mathbf{P}(H_1)q_1^{k^*} \\
&= \frac{1}{2}(1 + q_0^{k^*} - q_1^{k^*})
\end{aligned}$$

To find the value of k^* that minimizes p_e, we temporarily treat k^* as a continuous variable and differentiate p_e with respect to k^*. Setting this derivative to zero, we obtain

$$\frac{dp_e}{dk^*} = \frac{1}{2}((\log q_0)q_0^{k^*} - (\log q_1)q_1^{k^*}) = 0.$$

The solution to this equation is

$$\bar{k} = \frac{\log(|\log q_0|) - \log(|\log q_1|)}{|\log q_0| - |\log q_1|}.$$

As k^* ranges from 0 to $\bar{k}$, the derivative of p_e is nonzero, so that p_e is monotonic. Since $q_1 > q_0$, the derivative is negative at $k^* = 0$. This implies that p_e is monotonically decreasing as k^* ranges from 0 to $\bar{k}$. Similarly, the derivative of p_e is positive for very large values of k^*, which implies that p_e is monotonically increasing as k^* ranges from $\bar{k}$ to infinity. It follows that $\bar{k}$ minimizes p_e. However, k^* can only take integer values, so the integer k^* that minimizes p_e is either $\lfloor \bar{k} \rfloor$ or $\lceil \bar{k} \rceil$, whichever gives the lower value of P_e.

We now derive the form of the MAP decision rule, which minimizes the probability of error, and show that it is of the same type as the decision rules we just studied. With the MAP decision rule, for any given k, we accept H_1 if

$$\mathbf{P}(K = k \mid H_1)\mathbf{P}(H_1) > \mathbf{P}(K = k \mid H_0)\mathbf{P}(H_0),$$

and accept H_0 otherwise. Note that if

$$(1-q_1)q_1^k\mathbf{P}(H_1) > (1-q_0)q_0^k\mathbf{P}(H_0),$$

then

$$(1-q_1)q_1^{k+1}\mathbf{P}(H_1) > (1-q_0)q_0^{k+1}\mathbf{P}(H_0),$$

since $q_1 > q_0$. Similarly, if

$$(1 - q_1)q_1^k \mathbf{P}(H_1) < (1 - q_0)q_0^k \mathbf{P}(H_0),$$

then

$$(1 - q_1)q_1^{k-1} \mathbf{P}(H_1) < (1 - q_0)q_0^{k-1} \mathbf{P}(H_0).$$

This implies that if we decide in favor of H_1 when a value k is observed, then we also decide in favor of H_1 when a larger value is observed. Similarly, if we decide in favor H_0 when a value k is observed, then we also decide in favor of H_0 when a smaller value k is observed. Therefore, the MAP rule is of the type considered and optimized earlier, and thus will not result in a lower value of p_e.

(c) As in part (b) we have

$$p_e = \mathbf{P}(H_1) + \mathbf{P}(H_0)q_0^{k^*} - \mathbf{P}(H_1)q_1^{k^*}.$$

Consider the case where $\mathbf{P}(H_1) = 0.7$, $q_0 = 0.3$ and $q_1 = 0.7$. Using the calculations in part (b), we have

$$\bar{k} = \frac{\log\left(\frac{\mathbf{P}(H_0)\log(v_0)}{\mathbf{P}(H_1)\log(v_1)}\right)}{\log\left(\frac{v_1}{v_0}\right)} \approx 0.43$$

Thus, the optimal value of k^* is either $\lfloor \bar{k} \rfloor = 0$ or $\lceil \bar{k} \rceil = 1$. We find that with either choice the probability of error p_e is the same and equal to 0.3. Thus, either choice minimizes the probability of error.

Note that $\bar{k}$ decreases as $\mathbf{P}(H_1)$ increases from 0.7 to 1.0. So the choice $k^* = 0$ remains optimal in this range. As a result, we always decide in favor of H_1, and the probability of error is $p_e = \mathbf{P}(H_0) = 1 - \mathbf{P}(H_1)$.