Answers

1. Let

$$\begin{aligned}
S &= \text{the patient has strep throat} \\
\neg S &= \text{the patient does not have strep throat} \\
Y &= \text{a positive test result} \\
N &= \text{a negative test result}
\end{aligned}$$

   The prior is $\Pr(S) = 0.6$. The lab test has the following properties:

$$\begin{array}{ll}
\Pr(Y|S) = 0.7 & \Pr(N|S) = 0.3 \\
\Pr(Y|\neg S) = 0.1 & \Pr(N|\neg S) = 0.9.
\end{array}$$

   We need to find $\Pr(S|YNYNY)$.

$$\begin{aligned}
\Pr(YNYNY|S) &= 0.7 \cdot 0.3 \cdot 0.7 \cdot 0.3 \cdot 0.7 = 0.03087 \\
\Pr(YNYNY|\neg S) &= 0.1 \cdot 0.9 \cdot 0.1 \cdot 0.9 \cdot 0.1 = 0.00081.
\end{aligned}$$

   From Bayes rule, we have

$$\begin{aligned}
&\Pr(S|YNYNY) \\
&\quad = \frac{\Pr(S)\Pr(YNYNY|S)}{\Pr(S)\Pr(YNYNY|S) + \Pr(\neg S)\Pr(YNYNY|\neg S)} \\
&\quad = \frac{0.6 \cdot 0.03087}{0.6 \cdot 0.03087 + 0.4 \cdot 0.00081} \\
&\quad = 0.018522/0.018846 = 0.98
\end{aligned}$$

2. The proof can be found on pages 670-672 of the textbook. Two steps are required. We first need to establish that it is always possible to find a decision list consistent with any given data. A simple greedy algorithm that iteratively picks out pure subsets of the data works.

   We know from Section 18.5 of the textbook that a learning algorithm that always return a consistent hypothesis requires

$$N \geq \frac{1}{\epsilon}\left(\ln\frac{1}{\delta} + \ln|H|\right) \tag{1}$$

   examples to achieve error lower than ϵ with probability $1-\delta$. It is now a straightforward matter of working out $|k\text{-}DL(n)|$ and plugging that into (1) to see that the number of samples needed is a polynomial in the relevant parameters.

3. This is a reading exercise.

4. The ensemble makes an error if more than $\ulcorner M/2 \urcorner$ hypotheses in the ensemble makes an error. This happens with probability

$$\sum_{i=\ulcorner M/2 \urcorner}^{M} \binom{M}{i} \epsilon^i (1-\epsilon)^{M-i}.$$

For example, given $M = 5$ and $\epsilon = 0.1$, the error probability of the ensemble is

$$\binom{5}{3} 0.1^3 0.9^2 + \binom{5}{4} 0.1^4 0.9 + \binom{5}{5} 0.1^5 = 0.00856$$

5. The following is the derivation from the relevant slide.

$$\begin{aligned}
&\Pr(B|j,m) \\
&= \alpha \Pr(B) \sum_e \Pr(e) \sum_a \Pr(a|B,e) \Pr(j|a) \Pr(m|a) \\
&= \alpha \Pr(B) \sum_e \Pr(e) \sum_a f_A(a,b,e) f_J(a) f_M(a) \\
&= \alpha \Pr(B) \sum_e \Pr(e) f_{\bar{A}JM}(b,e) \quad \text{(sum out A)} \\
&= \alpha \Pr(B) f_{\bar{E}\bar{A}JM}(b) \quad \text{(sum out E)} \\
&= \alpha f_B(b) \times f_{\bar{E}\bar{A}JM}(b)
\end{aligned}$$

The different factors are given below.

$$\begin{aligned}
f_M(a) &= \mathit{if}\ a = \top\ \mathit{then}\ 0.7\ \mathit{else}\ 0.01 \\
f_J(a) &= \mathit{if}\ a = \top\ \mathit{then}\ 0.9\ \mathit{else}\ 0.05
\end{aligned}$$

$$\begin{aligned}
&f_A(a,b,e) = \\
&\quad \mathit{if}\ b = \top \wedge e = \top\ \mathit{then}\ (\mathit{if}\ a = \top\ \mathit{then}\ 0.95\ \mathit{else}\ 0.05) \\
&\quad \mathit{else\ if}\ b = \top \wedge e = \bot\ \mathit{then}\ (\mathit{if}\ a = \top\ \mathit{then}\ 0.94\ \mathit{else}\ 0.06) \\
&\quad \mathit{else\ if}\ b = \bot \wedge e = \top\ \mathit{then}\ (\mathit{if}\ a = \top\ \mathit{then}\ 0.29\ \mathit{else}\ 0.71) \\
&\quad \mathit{else}\ (\mathit{if}\ a = \top\ \mathit{then}\ 0.001\ \mathit{else}\ 0.999)
\end{aligned}$$

$$
\begin{aligned}
&f_{\bar{A}JM}(b, e) \\
&\quad = f_A(\top, b, e) f_J(\top) f_M(\top) + f_A(\bot, b, e) f_J(\bot) f_M(\bot) \\
&\quad = (\textit{if } b = \top \wedge e = \top \textit{ then } 0.95 \\
&\qquad \textit{else if } b = \top \wedge e = \bot \textit{ then } 0.94 \\
&\qquad \textit{else if } b = \bot \wedge e = \top \textit{ then } 0.29 \\
&\qquad \textit{else } 0.001) \cdot 0.9 \cdot 0.7 \\
&\qquad + \\
&\qquad (\textit{if } b = \top \wedge e = \top \textit{ then } 0.05 \\
&\qquad \textit{else if } b = \top \wedge e = \bot \textit{ then } 0.06 \\
&\qquad \textit{else if } b = \bot \wedge e = \top \textit{ then } 0.71 \\
&\qquad \textit{else } 0.999) \cdot 0.05 \cdot 0.01 \\
&\quad = \textit{if } b = \top \wedge e = \top \textit{ then } 0.60 \\
&\qquad \textit{else if } b = \top \wedge e = \bot \textit{ then } 0.59 \\
&\qquad \textit{else if } b = \bot \wedge e = \top \textit{ then } 0.18 \\
&\qquad \textit{else } 0.001
\end{aligned}
$$

$$
\begin{aligned}
&f_{\bar{E}\bar{A}JM}(b) = \\
&\quad = \Pr(e = \top) f_{\bar{A}JM}(b, \top) + \Pr(e = \bot) f_{\bar{A}JM}(b, \bot) \\
&\quad = 0.002 \cdot (\textit{if } b = \top \textit{ then } 0.6 \textit{ else } 0.18) \\
&\qquad + 0.998 \cdot (\textit{if } b = \top \textit{ then } 0.59 \textit{ else } 0.001) \\
&\quad = \textit{if } b = \top \textit{ then } 0.59 \textit{ else } 0.0014
\end{aligned}
$$

$$
f_B(b) = \textit{if } b = \top \textit{ then } 0.001 \textit{ else } 0.999
$$

6. It's easy to see that given the model M of a prediction suffix tree, $\Pr(x_{1:t}|M) = \prod_{l \in L(M)} \Pr_{kt}(x_{1:t|l}|M)$, where $L(M)$ denotes the leave nodes of M.

   Proposition 1. *Let D be the depth of the context tree. For each node n in the context tree at depth d, we have*

   $$P_w^n(x_{1:t|n}) = \sum_{M \in \mathcal{M}_{D-d}} 2^{-T_{D-d}(M)} P(x_{1:t|n}|M).$$

Proof. The proof proceeds by induction on d. The statement is clearly true for the leaf nodes at depth D. Assume now the statement is true for all nodes at depth $d+1$, where $0 \leq d < D$. Consider a node n at depth d. We have

$$\begin{aligned}
&P_w^n(x_{1:t|n}) \\
&= \frac{1}{2}P_{kt}(x_{1:t|n}) + \frac{1}{2}P_w^{n_l}(x_{1:t|n_l})P_w^{n_r}(x_{1:t|n_r}) \\
&= \frac{1}{2}P_{kt}(x_{1:t|n}) + \frac{1}{2}\Bigg(\sum_{M \in \mathcal{M}_{D-d-1}} 2^{-T_{D-d-1}(M)}P(x_{1:t|n_l}|M)\Bigg) \\
&\qquad\qquad\qquad\qquad\qquad \Bigg(\sum_{M \in \mathcal{M}_{D-d-1}} 2^{-T_{D-d-1}(M)}P(x_{1:t|n_r}|M)\Bigg) \\
&= \frac{1}{2}P_{kt}(x_{1:t|n}) + \\
&\quad \sum_{M_1 \in \mathcal{M}_{D-d-1}} \sum_{M_2 \in \mathcal{M}_{D-d-1}} 2^{-(T_{D-d-1}(M_1)+T_{D-d-1}(M_2)+1)}P(x_{1:t|n_l}|M_1)P(x_{1:t|n_r}|M_2) \\
&= \frac{1}{2}P_{kt}(x_{1:t|n}) + \sum_{(M_1,M_2) \in \mathcal{M}_{D-d}} 2^{-T_{D-d}((M_1,M_2))}P(x_{1:t|n} \mid (M_1, M_2)) \\
&= \sum_{M \in \mathcal{M}_{D-d}} 2^{-T_{D-d}(M)}P(x_{1:t|n} \mid M),
\end{aligned}$$

where (M_1, M_2) denotes the tree in $\mathcal{M}_{D-d}$ whose left and right subtrees are M_1 and M_2 respectively. $\square$

There are plenty of material on arithmetic coding on the web.