Problem 1

$$X = (age = youth, income = low, student = yes, credit_rating = fair)$$

From the table, $p(yes) = \frac{9}{14}$, and $p(no) = \frac{5}{14}$

We can also calculate:

$p(youth|yes) = \frac{2}{9},\ p(youth|no) = \frac{3}{5}$

$p(low|yes) = \frac{3}{9} = \frac{1}{3},\ p(low|no) = \frac{1}{5}$

$p(student = yes|yes) = \frac{6}{9} = \frac{2}{3},\ p(student = yes|no) = \frac{1}{5}$

$p(fair|yes = \frac{6}{9} = \frac{2}{3},\ p(fair|no) = \frac{2}{5}$

$$p(X|yes) = \frac{2}{9} \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} = 0.0329$$

$$p(X|no) = \frac{3}{5} \times \frac{1}{5} \times \frac{1}{5} \times \frac{2}{5} = 0.0096$$

Based on the computations above, we can see that the classifier would select yes - buys computer - for the new data X.

Problem 2

a) $t = 1$ ($S1$) is rain, and the sequence is $\{S1, S2, S3, S3, S1\}$

$$T = \begin{bmatrix} .4 & .3 & .3 \\ .2 & .6 & .2 \\ .1 & .1 & .8 \end{bmatrix}$$

$$\begin{aligned} p(\{S1, S2, S3, S3, S1\}) &= p(S1|S3)p(S3|S3)p(S2|S1)\pi_1 \\ &= (.1)(.8)(.2)(.3)(1) = .0048 \end{aligned}$$

b) Suppose the model is a known state S_α, then we know it will stay in that state for exactly d days, after which (at time $t = d + 1$), we know that it must transition to a different state. We can show the sequence for this as follows: $\{S_\alpha, S_\alpha, ..., S_\alpha, S \neq S_\alpha\}$ where the last state occurs at time $t = d + 1$. Now, we can compute the probability that this sequence will occur, knowing that the state will be S_α for exactly $1...d$ days:

$$\text{prob. of sequence } = p(S_\alpha) \times p(S_\alpha|S_\alpha) \times ... \times p(S_\alpha|S_\alpha) \times p(S \neq S_\alpha|S_\alpha)$$

$$\implies \text{prob. of sequence} = p(S_\alpha|S_\alpha)^{d-1} p(S \neq S_\alpha)$$

Which can be rewritten as elements in the transition matrix: prob of sequence $= (T_{\alpha\alpha})^{d-1} p(S \neq S_\alpha)$

$$\implies p(sequence) = (T_{\alpha\alpha})^{d-1}(1 - T_{\alpha\alpha})$$

c) From our answer above, we know that the probability of a sequence staying in a given state for d days is $(T_{\alpha\alpha})^{d-1}(1 - T_{\alpha\alpha})$, and the expectation value for the amount of observations of the sequence staying in the same consecutive state S_α would be given by:

$$\sum_{d=1}^{\infty} d(T_{\alpha\alpha})^{d-1}(1 - T_{\alpha\alpha})$$

Since we know that, at each iteration, there will be a $1 - T_{\alpha\alpha}$ probability of transitioning to a different state from our initial state S_α, and otherwise the probability of staying in our initial state would still be $T_{\alpha\alpha}$. Now, solving the summation based on the formula provided in the homework:

$$= (1 - T_{\alpha\alpha}) \frac{1}{(T_{\alpha\alpha} - 1)^2}$$

$$\implies \text{expected number of observations } = \frac{1}{1 - T_{\alpha\alpha}}$$

Problem 3

$$\text{state transition matrix } = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$$

$$\text{observation matrix } = \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{bmatrix}$$

The initial state is: $\pi = (0.6, 0.4)$
The observations sequence is: $O = \{S, L, L\}$ From the observation sequence, we can compute all of the probabilities of a state with sequence size 3 through a brute force strategy:
i.e. $p(HHH) = \pi_H p(H|S)p(H|H)p(H|L)p(H|H)p(H|L)$ etc...

$p(HHH)$	$(.6)(.1)(.7)(.5)(.7)(.5) = 0.00735$
$p(CCC)$	$(.4)(.7)(.6)(.1)(.6)(.1) = 0.001008$
$p(HHC)$	$(.6)(.1)(.7)(.5)(.4)(.1) = 0.00084$
$p(HCH)$	$(.6)(.1)(.4)(.1)(.3)(.5) = 0.00036$
$p(CHH)$	$(.4)(.7)(.3)(.5)(.7)(.5) = 0.0147$
$p(CCH)$	$(.4)(.7)(.6)(.1)(.3)(.5) = 0.00252$
$p(HCC)$	$(.6)(.1)(.4)(.1)(.6)(.1) = 0.000144$
$p(CHC)$	$(.4)(.7)(.3)(.5)(.4)(.1) = 0.00168$

Now that we have found the overall distribution of all possible states with sequence size 3, we can determine the most likely sequence for the given observations. Note - these probabilities are not standardized.
Just by looking at the table, it seems that the most likely sequence in our case would be $\{C, H, H\}$. However, we can determine what the most likely sequence will be by seeing based on the table, what the most probable transition will be at each time step. Let's look at the standardized probabilities (computed by taking the sum of the probabilities in the table and dividing each element into it):

$p(HHH)$	0.257
$p(CCC)$	0.035
$p(HHC)$	0.029
$p(HCH)$	0.013
$p(CHH)$	0.514
$p(CCH)$	0.088
$p(HCC)$	0.005
$p(CHC)$	0.059

We can also verify that these probabilities will sum to 1 to ensure they are standardized correctly (they are). Thus, we can see that, for time step t_0, we have $p(H) = 0.257+0.029+0.013+0.005 = 0.304$, and so we know $p(C) = 1 - p(H) = 0.696$. Next, for t_1, $p(H) = 0.257 + 0.029 + 0.514 + 0.059 = 0.859$ and $p(C) = 1-p(H) = 0.141$. Finally, for t_2, $p(H) = 0.257+0.013+0.514+0.088 = 0.872$ and $p(C) = 0.128$. As we can see, this matches our prediction of $\{C, H, H\}$ as being the most likely sequence to occur.