CS 530 — Principles of AI
Sample Problems — Selected Answers (no pictures)

Problem 1a. $P(C)$, $P(J|C)$, $P(X|J)$, $P(Y|J)$.
Problem 1b. J takes values in 1..4.
$P(C=1) = P(C=2) = .5$
$P(J=1|C=1) = P(J=2|C=1) = P(J=3|C=2) = P(J=4|C=2) = .5$
(all other values for $P(J|C) = 0$).
$P(X = \text{true}|J=1) = P(Y = \text{false}|J=1) = 1.$
$P(X = \text{false}|J=2) = P(Y = \text{true}|J=2) = 1.$
$P(X = \text{true}|J=3) = P(Y = \text{true}|J=3) = 1.$
$P(X = \text{false}|J=4) = P(Y = \text{false}|J=4) = 1.$
(all other values for $P(X|J)$ and $P(Y|J)$ are 0)
Problem 1c. No: X and Y are not conditionally independent given C.
Problem 1d. Yes. This doesn't make any independence assumptions: it's just a factorization in terms of the chain rule $P(C)P(X|C)P(Y|C,X)$.
Problem 4. $P(x|C=c1)P(C=c1) = 3/4 * 1/3 = 1/4$
$P(x|C=c2)P(C=c2) = 1/4 * 2/3 = 1/6$
$P(C=c1|x) = (1/4)/(1/4+1/6) = 3/5$
Problem 5. $U(s,c1) = 1$, $U(s,c2) = 0$.
$U(d,c1) = .4$, $U(d,c2) = .4$.
$U(h,c1) = 1-q$, $U(h,c2) = .4-q$.
Problem 6. Assume $P(C=c1) = .6$. Then $U(s) = .6$. $U(d) = .4$. $U(h) = .76-q$. So h is best when $q < .16$.
Problem 8. Define a feature corresponding to presence of each subsequence. Model feature presence as a function of category, for example using a naive Bayes model.
Problem 9. You would need many states in an HMM, in order to track the presence of suitable subsequences. This would lead to sparse data.
Problem 10. Set up the HMM so that it has two underlying states, and you can transition freely (but rarely) between them. Expect HMM learning to cluster these states appropriately.
Problem 11. The only way to get from the sequence to features is by tracking the distribution of symbols in the sequence per category. You need to learn this from data. But a modeling method like a naive Bayes model requires you to predefine the observations you make before learning.
Problem 12. Pick the A that maximizes $\sum_D U(A,D)P(D|C)$.
Problem 14. $P(C), P(B), P(X|C), P(Y|C)$.
Problem 15. $p(C), p(B), p(X|C), p(Y|C)$.
Problem 16. Classify into the C that maximizes $P(X,Y,C)$.
In other words, classify into the C that maximizes

$$\sum_B P(X|B,C)P(Y|B,C)P(B)P(C)$$

Problem 17. Estimate $\hat{C}$ to maximize

$$\int p(X|B,C)p(Y|B,C)p(B)p(C)dB$$

Problem 18. Take (16). As B is known better, the summation

$$\sum_B P(X|B,C)P(Y|B,C)P(B)P(C)$$

will be increasingly dominated by the specific values of B for which $P(B)$ is high. This means we zero in on particular entries in the distributions for $P(X|B,C)$ and $P(Y|B,C)$, and so we have more constrained correlations between X and Y and specific values of C. This reduces the uncertainty in our estimate of C.

Problem 18. Create a range of versions of the system: a simple baseline (say, predict the most likely action given just one feature); the original version; a version that's sensitive to a few words; and a version that's sensitive to all words. Perhaps have a person who will watch transcripts and guess the user's action.

Collect new data for training and test—Make sure that you don't bias the evaluation by using your experimental data to design the system that uses few words. Use crossvalidation to get a sense of the statistical performance of the different systems. Make sure that performance differences are significant, and that baseline and top-performing systems are clearly different from other possibilities.