Book questions
2.4 Bayes rule for medical diagnosis,
2.5 Monty Hall problem
2.6 Conditional Independence
3.19 Irrelevant features with naive Bayes
3.22 Fitting a naive Bayes spam filter by hand

Additional questions:

1. Which of the following problems are more suited for a learning approach and which are more suited for a design approach? (from AML book)
   1. Determining the age at which a particular medical test should be performed
   2. Classifying numbers into primes and non-primes
   3. Detecting potential fraud in credit card charges
   4. Determining the time it would take a falling object to hit the ground
   5. Determining the optimal cycle for traffic lights in a busy intersection

2. You have an unfair 6-sided die (i.e. 1 dice). Values 1 and 4 are rolled with probability 1/4. All other values are rolled with probability 1/8.
   (a) What is the formula for expected value in the discrete case?
   (b) What is the expected value of a single throw of the biased die?
   (c) What is the variance of the biased die?

3. Assume we know that $x_1, x_2, \ldots, x_n \sim N(\mu, \sigma^2)$ but we don’t know the value of $\mu$. Construct and solve for a likelihood function such that we maximize $P(x|\mu, \sigma)$ [Hint: you will need to know the analytical form of the normal equation, also assume $x$ is iid. Hint 2: Computing the log likelihood is easiest.]

4. Bias-variance decomposition (from ISLR).
   (a) Provide a sketch of typical (squared) bias, variance, training error, test error, and Bayes (or irreducible) error curves, on a single plot, as we go from less flexible statistical learning methods towards more flexible approaches. The x-axis should represent the amount of flexibility in the method, and the y-axis should represent the values for each curve. There should be five curves. Make sure to label each one.
   (b) Explain why each of the five curves has the shape displayed in part (a).
   (c) Add labels to your plot indicating regions of over- and under-fitting. Explain the reason for over or underfitting of the testing data in terms of bias and variance.