Administrative comments: The following are designed to be questions that could appear on a midterm in our course. Many of the questions are based on books referred to during our course such as Cormen et al.

1. Matlab question: There are different types of Matlab questions that could appear. One could be to write a few lines of code. Another is where I insert errors into some Matlab code and ask you to spot several of them (for each one, need to explain how to correct it). For example, the code below contains 4 mistakes; typically I would insert 7-8 mistakes and ask to spot roughly half.

```matlab
function y=mergesort(x)
    if length(x)<3 % x is short --> already sorted --> return it
        x=y;
    else % x needs to be sorted
        N=length(x);
        N2=N/2;
        y1=mergesort(x(1:N2)); % sort first half
        y2=mergesort(x(N2+1:N)); % second half
        y=merge(x1,x2);
    end
end
```

2. Consider a model for text, where the alphabet has C characters, and each character is predicted by considering the previous K characters. (That is, \(X_{n+k}\) is governed by a probability mass function, \(Pr(X_{n+k}|X_n,X_{n+1},\ldots,X_{n+k-1})\).) We want to learn these probabilities from the data. Please describe the model complexity for length-N input strings as a function of C, K, and N.

3. Derive a Bayesian classification approach for two classes governed by the following distributions, where each data point lies in \(p=1\)-dimensional space, \(f_{\text{blue}}=0.5N(-1,1)+0.5N(2,1)\) and \(f_{\text{red}}=N(0,1)\), where the Gaussian means of the three components are different, the variances of all Gaussian components is 1, and the 0.5 values in the blue class correspond to the probabilities of the two Gaussian components.

   In particular, given \(X \in \mathbb{R}^p\), describe the posterior probability \(Pr(\text{red}|X)\). When invoking Bayes’ rule, you can assume that \(Pr(\text{blue})=Pr(\text{red})=0.5\).

4. Repeat occurrences: Consider the problem of determining whether a sequence of \(n\) numbers contains \(n\) distinct numbers, or instead at least one number occurs multiple times. Provide a \(\Theta(n \times \log(n))\) algorithm for doing so.
5. Express $0.001n^3+100n^2$ in terms of $\Theta$ notation.

6. Suppose that merge sort runs in $64n \times \log_2(n)$ steps while insertion sort takes $8n^2$.
   a. For which value of $n$ does merge sort start beating insertion sort?
   b. How might you modify merge sort to obtain faster performance on small inputs? Discuss the modification and resulting running time.

7. Show that in an undirected graph the length of a cycle must be at least 3.