Please typeset your answers, or write very clearly. If TAs cannot read your handwriting, they will not grade your assignment.

1. (20 pts) Prove the following two properties of the Huffman encoding scheme.
   - If some character occurs with frequency strictly more than $2/5$, then there is guaranteed to be a codeword of length 1.
   - If all the characters occur with frequency strictly less than $1/3$, then no codeword of length 1 will be produced.

2. (15 pts) You want to typeset a textfile by determining where to add line breaks. Suppose the $i$th word in the file has length $w_i$ (that is, it takes up $w_i$ character spaces), where for simplicity we assume that no “blank spaces” are required between the words.

   A finicky editor has decided that the ideal line length is $L$. That is, a line with length $K$ incurs a penalty of $L - K$. (No line can be longer than $L$, so the penalty $L - K$ is always non-negative.)

   We want to minimize the total penalty of a page layout, which is defined as the sum of individual line penalties. Design a greedy algorithm for minimizing the total penalty. First concisely describe your algorithm, then prove its correctness.

3. (15 pts) Reconsider the page layout problem of Problem 2 but now we want to minimize the maximum penalty of any line in the page, not the total penalty. (For example, if in a layout, there are three lines, with penalties 10, 6 and 8. Then, the penalty for the page layout is 10.) We want to design the layout so that the page penalty is smallest among all layouts. Does your greedy algorithm solve this problem optimally as well? Prove or disprove.