Comparison Based Merging

Upper and Lower bounds
Merging

**Input:** Two sorted lists $X$ and $Y$ of length $n$ and $m$.

We may assume $n \geq m$.

\[ n \]

$X$: ________________________________

\[ m \]

$Y$: __________________

**Theorem:**

In a comparison based model, the complexity of merging $X$ and $Y$ is

\[ \Theta(m(\log(n/m) + 1)) \]
Simple Upper Bounds

Standard Merge:  
\[ \Theta(n + m) \]

Binary Insertion of \( Y \) in \( X \):  
\[ \Theta(m \log n) \]

For "large" \( m \) (\( m = \Theta(n) \)):  
\[ \Theta(n + m) = \Theta(m(\log(n/m) + 1)) \]

For "small" \( m \) (e.g. \( m = O(\sqrt{n}) \)):  
\[ \Theta(m \log n) = \Theta(m(\log(n/m) + 1)) \]
The Simple Bounds are Sub-Optimal

E.g. for $m = \Theta(n / \log n)$:

$$\Theta(n + m) = \Theta(n)$$

$$\Theta(m \log n) = \Theta(n)$$

$$\Theta(m(\log(n/m) + 1)) = \Theta(n \frac{\log \log n}{\log n}) = o(n)$$
Graphically

\[ n + m = 200 \]

\[ m \log n \]

\[ m(\log(n/m) + 1) \]
Better Upper Bound

\[ \text{if } x_i < y_j \]
\[ i++ \]
\[ \text{else} \]
\[ \text{binary search from } x_{i-1} \text{ to } x_i \]
\[ j++ \]

Number of comparisons:

\[ m + m \log(n/m) \]
Lower Bound

There are \( \binom{n+m}{m} \) different possible results of the merging two sorted lists of lengths \( n \) and \( m \).

So any decision tree for merging must have at least that many leaves.

It must hence have height at least

\[
\log\left(\binom{n+m}{m}\right)
\]
Lemmas

For $n \geq m$:

1)  
\[
\binom{n + m}{m} = \frac{(n + m)(n + m - 1) \cdots (n + 1)}{m(m - 1) \cdots 1} \geq \left( \frac{n}{m} \right)^m
\]

2)  
\[
\binom{n + m}{m} \geq \binom{2m}{m} \geq \frac{2m(2m - 1) \cdots (m + 1)}{m(m - 1) \cdots 1}
\]

\[
\geq 2\left( \frac{m}{m} \right) 2\left( \frac{m - 1/2}{m - 1} \right) 2\left( \frac{m - 2/2}{m - 2} \right) 2\left( \frac{m - 3/2}{m - 3} \right) \cdots \geq 2^m
\]
Lemmas

3) 

\[ h(n) \geq f(n) \text{ and } h(n) \geq g(n) \]

\[ \iff \]

\[ h(n) \geq \max\{f(n), g(n)\} \]

4) 

For \( f \) and \( g \) positive:

\[ \max\{f(n), g(n)\} = \Theta(f(n) + g(n)) \]
Lower Bound Computation

\[
\log\left(\binom{n + m}{m}\right)
\geq \max\{\log(2^m), \log((n/m)^m)\}
= \max\{m, m \log(n/m)\}
= \Omega(m + m \log(n/m))
\]