Dimostrazione scelta greedy per Huffman

$$B(T) = \sum_{c \in C} c.freq \cdot d_T(c)$$

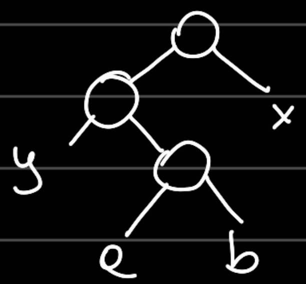
costo totale — frequenza — bit di codifica

Siano x e $y \in C$ con le frequenze minime. Allora esiste un codice prefix-free ottimo in cui le codifiche di x e y hanno la stessa lunghezza massima e differiscono per un solo bit l'ultimo

DIMOSTRAZIONE

T è l'albero che rappresenta il codice

T :



a e b sono foglie sorelle di massima profondità

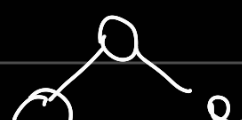
Avremo che $x.freq \leq a.freq$ e $y.freq \leq b.freq$

Può accadere $x.freq = a.freq$ o $y.freq = b.freq$

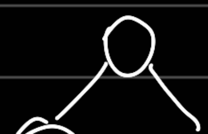
Se $x.freq = b.freq$ il lemma risulta immediatamente vero

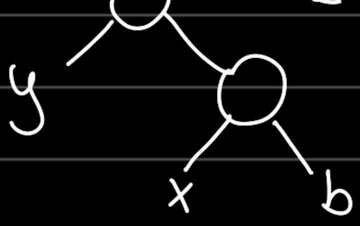
Se $x.freq < b.freq$ creiamo:

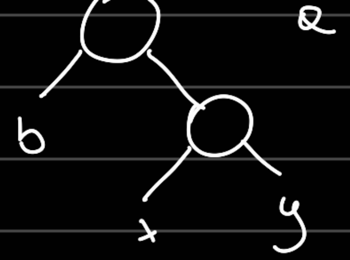
T'



T''







Dobbiamo dimostrare che $B(T'') \leq B(T') \leq B(T)$
per farlo dimostriamo che $B(T) - B(T') \geq 0$

$$= \underbrace{\sum_{c \in C - \{a,x\}} c.freq \cdot d_T(c) - \sum_{c \in C - \{a,x\}} c.freq \cdot d_{T'}(c)}_{=0} + x.freq\left(d_T(x) - d_{T'}(x)\right) + a.freq\left(d_T(x) - d_{T'}(x)\right)$$

$$= x.freq\left(d_T(x) - d_T(a)\right) + a.freq\left(d_T(a) - d_T(x)\right)$$

$$= \underbrace{(a.freq - x.freq)}_{\geq 0}\ \underbrace{(d_T(a) - d_T(x))}_{\geq 0} \geq 0$$

Si fa analogamente per T'' ed essendo T un albero ottimo si verifica che tutte le configurazioni sono uguali

LEMMA: Siano $x, y \in C$ di freq. minima
Sia $C' = (C - \{x, y\}) \cup \{z\}$. Definiamo $C.freq$ in C: $C.freq$ in C' e $z.freq = x.freq + y.freq$
Sia T' un albero che rappresenta il codice prefix-free ottimo per C'. Sia T ottenuto rimpiazzando z con un nodo genitore di x e y
Allora T è la rappresentazione di un codice prefix-free ottimo per C

DIMOSTRAZIONE:

$$\forall c \in C - \{x, y\} : d_T(c) = d_{T'}(c) \Rightarrow$$

$$c.freq \cdot d_T(c) = c.freq \cdot d_{T'}(c)$$

Poiché $d_T(x) = d_T(y) = d_T(z) + 1 \Rightarrow$

$$x.freq \cdot d_T(x) + y.freq \cdot d_T(y) = (x.freq + y.freq)(d_T(z) + 1)$$

Allora:

$$B(T) = \sum_{c \in C / \{x, y\}} c.freq \cdot d_T(c) + (x.freq + y.freq) \cdot (d_T(z) + 1)$$

$$B(T) = B(T') + z.freq$$

$$B(T') = B(T) - (x.freq + y.freq)$$

Se T non fosse un albero ottimo per $C \Rightarrow \exists\, T''$ per C:

$$B(T'') < B(T)$$

T''' è ottenuto da T'' rimpiazzando x e y e il loro genitore con una foglia z tale che $z.freq = x.freq + y.freq$

Si verifica che si va a creare una contraddizione con l'ipotesi che T' è un albero ottimo per C