1. Suponga que cuenta con los caracteres a,b,c,d,e,f tienen probabilidades según la siguiente matriz. Encuentre el árbol de huffman correspondiente, calculando la probabilidad de cada letra que se encuentra en la matriz **(15 pts)**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| a | c | e | e | e | e | d | d | d | d |
| b | a | e | e | e | e | b | b | b | d |
| f | b | a | c | c | c | a | d | b | d |
| f | c | b | a | c | c | a | d | b | d |
| f | c | b | b | a | c | b | a | d | d |
| f | c | e | a | b | c | b | d | a | e |
| f | a | a | f | b | b | b | d | d | a |
| f | f | f | f | b | c | c | d | d | d |
| a | f | f | a | d | a | a | f | f | f |
| f | a | a | f | a | a | d | a | a | f |

Nota: Probabilidad de “a” se calcula contando la cantidad de “a” y la cantidad total

Probabilidad=CA/TL -> probabilidad=22/100 ->0.22. La probabilidad de “a” seria 0.22.

Texto

Descripción generada automáticamente

Algoritmo de Huffman

El **algoritmo de Huffman** es un algoritmo para la construcción de códigos de Huffmann, desarrollado por David A. Huffmann en 1952 y descrito en “A Method for the Construction of Minimum-Redundancy Codes”.

Este algoritmo toma un alfabeto de *n* símbolos, junto con sus frecuencias de aparición asociadas, y produce un código de Huffman para ese alfabeto y esas frecuencias.

# Descripción

El algoritmo consiste en la creación de un árbol binario que tiene cada uno de los símbolos por hoja, y construido de tal forma que siguiéndolo desde la raíz a cada una de sus hojas se obtiene el código Huffman asociado.

* 1. Se crean varios árboles, uno por cada uno de los símbolos del alfabeto, consistiendo cada uno de los árboles en un nodo sin hijos, y etiquetado cada uno con su símbolo asociado y su frecuencia de aparición.
  2. Se toman los dos árboles de menor frecuencia, y se unen creando un nuevo árbol. La etiqueta de la raíz será la suma de las frecuencias de las raíces de los dos árboles que se unen, y cada uno de estos árboles será un hijo del nuevo árbol. También se etiquetan las dos ramas del nuevo árbol: con un 0 la de la izquierda, y con un 1 la de la derecha.
  3. Se repite el paso 2 hasta que sólo quede un árbol.

Con este árbol se puede conocer el código asociado a un símbolo, así como obtener el símbolo asociado a un determinado código.

Para obtener el código asociado a un símbolo se debe proceder del siguiente modo:

1. Comenzar con un código vacío
2. Iniciar el recorrido del árbol en la hoja asociada al símbolo
3. Comenzar un recorrido del árbol hacia arriba
4. Cada vez que se suba un nivel, añadir al código la etiqueta de la rama que se ha recorrido
5. Tras llegar a la raíz, invertir el código
6. El resultado es el código Huffman deseado

Para obtener un símbolo a partir de un código se debe hacer así:

1. Comenzar el recorrido del árbol en la raíz de éste
2. Extraer el primer símbolo del código a descodificar
3. Descender por la rama etiquetada con ese símbolo
4. Volver al paso 2 hasta que se llegue a una hoja, que será el símbolo asociado al código

En la práctica, casi siempre se utiliza el árbol para obtener todos los códigos de una sola vez; luego se guardan en tablas y se descarta el árbol.

# 1. Ejemplo de uso

La tabla describe el alfabeto a codificar, junto con las frecuencias de sus símbolos. En el gráfico se muestra el árbol construido a partir de este alfabeto siguiendo el algoritmo descrito.

Diagrama

Descripción generada automáticamente

Árbol para construir el código Huffman del ejemplo SímboloFrecuencia A0,15 B0,30 C0,20 D0,05 E0,15 F0,05 G0,10

Se puede ver con facilidad cuál es el código del símbolo **E**: subiendo por el árbol se recorren ramas etiquetadas con **1**, **1** y **0**; por lo tanto, el código es **011**. Para obtener el código de **D** se recorren las ramas **0**, **1**, **1** y **1**, por lo que el código es **1110**.

La operación inversa también es fácil de realizar: dado el código **10** se recorren desde la raíz las ramas **1** y **0**, obteniéndose el símbolo **C**. Para descodificar **010** se recorren las ramas **0**, **1** y **0**, obteniéndose el símbolo **A**.