

## Proyecto de Redes Bayesianas

### 1 primera parte

En este proyecto vamos a usar una red bayesiana para calcular probabilidades de diferentes escenarios en competiciones de deportes de equipo.

En esta primera parte empezaremos considerando una competición con un sistema de **eliminatória**.

Por lo tanto, tenemos  $n$  equipos ( $n$  es una potencia de 2) sobre los que con algún mecanismo prefijado (típicamente un sorteo) se fija una estructura de árbol binario tal como la que aparece en la figura siguiente.



En este tipo de competición no hay empates y el árbol fija el emparejamiento de los partidos en los que van jugando los equipos que van superando las diferentes etapas de la eliminatoria.

En la primera entrega de la práctica tienes que diseñar una Red Bayesiana para este tipo de competición. La entrega (fecha de entrega 13 de Noviembre) consistirá en una red bayesiana en formato SamIam para  $n = 4$  que permita hacer predicciones sobre los ganadores y que éstas se puedan ir actualizando tal como se vaya obteniendo evidencia de los resultados. También hay que entregar un breve documento que explique cómo sería la red para una  $n$  (potencia de dos) arbitraria explicando las variables, los arcos y las tablas de probabilidad (CPTs) que habría que poner. Esta entrega valdrá el 30% de la nota de la práctica. Se puede hacer por parejas.

**Solucion:** La Red tiene la misma estructura que el árbol de eliminatorias. Cada nodo representa quien ha ganado el partido correspondiente. Por lo tanto el dominio de cada variable se corresponde con los equipos que pueden haber llegado a ese punto (la variable del nodo de la final tiene a todos los equipos en su dominio). Explico a continuación la CPT del nodo de la final, porque la idea es la misma para todos los nodos. El nodo de la final tiene dominio  $1, 2, \dots, n$  y sus dos padres tienen dominio  $1, 2, \dots, n/2$  y  $(n/2) + 1, (n/2) + 2, \dots, n$ . La probabilidad  $P(a|b, c)$  será 0 si  $a \neq b$  y  $a \neq c$ . En caso contrario será el valor que consideremos en función de la calidad futbolística de  $b$  y  $c$ .

### 2 Segunda Parte

En la segunda parte consideraremos una competición con un sistema de **liguilla a una vuelta**.

Por lo tanto, tenemos  $n$  equipos ( $n$  es un número par) y cada jornada se emparejan de manera diferente de manera que tras  $n - 1$  jornadas todos los equipos han jugado contra todos los equipos, tal como se ve en la imagen para  $n = 8$

Round 1	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7
2 vs 1	3 vs 4	6 vs 2	7 vs 5	1 vs 3	4 vs 5	7 vs 3
3 vs 8	1 vs 7	7 vs 8	8 vs 4	4 vs 2	8 vs 1	8 vs 2
4 vs 7	8 vs 6	4 vs 1	2 vs 3	5 vs 8	2 vs 7	1 vs 5
5 vs 6	2 vs 5	5 vs 3	6 vs 1	6 vs 7	3 vs 6	6 vs 4

La decisión de quién juega contra quién en cada jornada se hace con algún mecanismo prefijado (típicamente un sorteo) pero nosotros asumiremos que esto ya viene dado. También asumiremos que no hay empates y que el resultado de la liguilla viene dado por el número de partidos que ha ganado cada equipo.

En la segunda entrega de la práctica tienes que diseñar una Red Bayesiana para este tipo de competición. La entrega (fecha de entrega 24 de Noviembre) consistirá en una red bayesiana en formato SamIam para  $n = 3$  que permita hacer predicciones sobre los ganadores y que éstas se puedan ir actualizando tal como se vaya obteniendo evidencia de los resultados. La red tiene que tener una variable Ganador (G) con cuatro valores que representen los cuatro equipos que pueden ser ganadores. Como puede pasar que varios equipos acaben la liga empatados a puntos, podeis usar cualquier criterio arbitrario para decidir quien gana entre ellos (por ejemplo, lexicograficamente). También hay que entregar un breve documento que explique cómo sería la red para una  $n$  arbitraria explicando las variables, los arcos y las tablas de probabilidad (CPTs) que habría que poner. Esta entrega valdrá el 40% de la nota de la práctica. Se puede hacer por parejas.

**Solución:** La solución más natural es tener una estructura matricial de nodos-partido y nodos-puntuación donde cada columna se corresponde a una jornada. En cada columna hay un nodo-partidos por cada partido que se juega en esa jornada (en el caso de  $n = 3$  sólo habrá uno por columna) y tienen dos valores posibles que se corresponden al posible ganador del partido. Los nodos-partido no tienen padre y su distribución de probabilidad dependerá de la calidad de los dos equipos. En cada columna hay un nodo-puntuación por equipo que representa los puntos acumulados de ese equipo (los nodos-puntuacion tienen dos padres, la puntuacion anterior de ese equipo y el nuevo partido de ese equipo). La CPT de los nodos-puntuación es determinista. A la derecha de la estructura matricial (es decir, después de la última jornada) hay que poner un nodo-ganador G con  $n$  valores posibles y cuyos padres son las puntuaciones de los  $n$  equipos tras la ultima jornada. Date cuenta que la CPT de esta variable G es determinista. Con  $n = 3$  la CPT es grande pero manejable. Para  $n$  mas grandes habría que usar una forma más compacta de representar la CPT (algo que no se puede hacer con SamIam)

### 3 Tercera Parte

En la tercera parte consideraremos una competición que funciona en dos fases. Compiten 8 equipos a los que llamaremos  $e_1, e_2, \dots, e_8$ . El equipo  $e_1$  es el equipo anfitrión y el  $e_8$  es el equipo que ganó la ultima edición de esta competición. En la primera fase de la competición dividimos 6 equipos en dos grupos:  $e_2, e_3, e_4$  y  $e_5, e_6, e_7$ . Cada grupo hace una liguilla tal como vimos en la segunda entrega. El anfitrión y el ganador de la última edición no hacen liguilla y pasan a la segunda fase directamente. La segunda fase es eliminatoria (tal como vimos en la primera entrega). En el primer nivel de la eliminatoria juega el anfitrión  $e_1$  con el ganador de la primera liguilla (es decir  $e_2$  o  $e_3$  o  $e_4$ ), y el ganador de la última edición  $e_8$  juega con el ganador de la segunda liguilla. Los ganadores de estos partidos juegan la final.

En la tercera entrega de la práctica tienes que diseñar una Red Bayesiana para esta competición a dos fases. La entrega (fecha de entrega 1 de Diciembre) consistirá en una red bayesiana en formato SamIam que permita hacer predicciones sobre los ganadores y que éstas se puedan ir actualizando tal como se vaya obteniendo evidencia de los resultados. La red tiene que tener una variable Ganador (G) con 8 valores que representen los 8 equipos que pueden ser ganadores. Notad que la CPT de la variable G va a ser bastante grande, aunque sólo 16 de sus entradas tendrán sentido.

En las entregas anteriores habíamos obviado el hecho de que en la vida real las probabilidades de que un equipo determinado gane a otro no son arbitrarias y tienen que reflejar de alguna manera nuestro conocimiento del problema (que pueden ser subjetivo u objetivo). En esta entrega supondremos que le hemos pedido a un experto en fútbol que nos haga un ranking de calidad de los 8 equipos y nos ha dado el siguiente orden (8, 4, 6, 1, 3, 2, 5, 7) (es decir el equipo 8 es el mejor y el 7 el peor). En esta entrega tienes que pensar en una forma de utilizar esta información para que las probabilidades de tu modelo sean más informadas.

Además de la red bayesiana en formato SamIam, teneis que entregar un pdf que contenga:

1. Imagen de la Red

2. Explicación del significado de las variables indicando claramente su dominio (es decir, los valores que puede tomar la variable)
3. Imagen de las CPTs (como hay CPTs repetidas, solo hace falta poner una de cada)
4. Explicación de cómo habéis usado la lista ordenada para calcular las probabilidades.
5. Imagen del monitor de la variable  $G$  sin ninguna evidencia.

Esta entrega valdrá el 30% de la nota de la práctica.