1. Pre-requisitos

1. Asegúrese de tener instalado Python en versión 3.7 o superior.

2. Para este taller usaremos Python y en particular la librería *pgmpy*. Instale *pgmpy* con el comando

   ```
   pip install pgmpy
   ```

   o si prefiere repositorios de anaconda use

   ```
   conda install -c ankurankan pgmpy
   ```

   También puede usar el administrador de paquetes de Anaconda.

3. *pgmpy* usa las siguientes librerías, así que es buena idea instalarlas previamente

   ```
   numpy
   scipy
   scikit-learn
   pandas
   pyparsing
   pytorch
   statsmodels
   tqdm
   joblib
   ```

4. La entrega de este taller consiste en un **reporte** y unos **archivos de soporte**. Cree el archivo de su **reporte** como un documento de texto en el que pueda fácilmente incorporar capturas de pantalla, textos y similares. Puede ser un archivo de word, libre office, markdown, entre otros.

2. El problema de Monty Hall

En un programa concurso Ud puede ganar un carro. Para hacerlo debe seleccionar una de tres puertas. Detrás de una de ellas está el carro y si la selecciona se gana el carro.

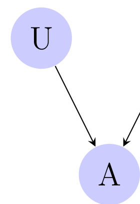
Tras su selección una de las puertas, el animador del programa abrirá otra de las puertas, detrás de la que hay una cabra. Recuerde que el animador sabe qué hay detrás de cada puerta.

Tras abrir la otra puerta, el animador le da la posibilidad de cambiar de puerta, ¿Ud se cambia o no de puerta?

1. Para decidir si debe cambiar o no de puerta Ud ha decidido modelar este problema como una Red Bayesiana.

2. Defina tres variables aleatorias U, C y A, que representan la puerta seleccionada por Ud, la puerta detrás de la cual está el carro, y la puerta que selecciona el animador, respectivamente. Las tres variables tienen espacio de estados $\{1, 2, 3\}$.

3. Ud ha definido la siguiente red bayesiana para modelar este problema. En su **reporte** explique por qué decide emplear esta red.



4. Responda las siguientes preguntas en su **reporte**:

   a) ¿Tiene esta red estructuras en V? Si las tiene, ¿Qué implicaciones tienen en la influencia entre las variables de la red?

   b) ¿Qué tipo de análisis considera es más relevante en este caso? ¿Causal, evidencial, intercausal? ¿Por qué?

   c) ¿Qué independencias condicionales captura esta red?

5. El carro ha sido ubicado al azar detrás de alguna de las puertas, y Ud selecciona una puerta al azar inicialmente. Modele este comportamiento con las distribuciones de probabilidad $P(U)$ y $P(C)$. **R**.

6. EL comportamiento del Animador (la puerta que decide mostrar) depende tanto de su selección como de la puerta donde se encuentra el carro. Modele el comportamiento del Animador con la distribución condicional de probabilidad $P(A|C, U)$ (note que esta representación es estocástica por columnas) **R**:

	C	1			2			3		
A	U	1	2	3	1	2	3	1	2	3
1		0								
2		0.5								
3		0.5								

7. En su editor de python cree un nuevo archivo monty.py e incluya los objetos BayesianNetwork y TabularCPD

```
from pgmpy.models import BayesianNetwork
from pgmpy.factors.discrete import TabularCPD
```

8. Defina ahora la estructura de la red incluyendo los arcos y nodos

```
model = BayesianNetwork([("C", "H"), ("P", "H")])
```

9. Defina las CPDs de C y U

```
cpd_c = TabularCPD(variable="C", variable_card=3, values=[[0.33],
    [0.33], [0.33]])
cpd_u = TabularCPD(variable="U", variable_card=3, values=[[0.33],
    [0.33], [0.33]])
```

10. Defina la CPD de A

```
cpd_a = TabularCPD(
    variable="A",
    variable_card=3,
    values=[
        [0, 0, 0, 0, 0.5, 1, 0, 1, 0.5],
        [0.5, 0, 1, 0, 0, 0, 1, 0, 0.5],
        [0.5, 1, 0, 1, 0.5, 0, 0, 0, 0],
    ],
    evidence=["C", "U"],
    evidence_card=[3, 3],
)
```

11. En la documentación de *pgmpy*podrá encontrar los detalles de la clase `TabularCPD`, ver `https://pgmpy.org/factors/discrete.html`. En su **reporte**justifique el valor de los argumentos usados al crear los tres objetos de la clase `TabularCPD`.

12. Asocie las 3 CPDs a su modelo

```
modelo.add_cpds(cpd_c, cpd_u, cpd_a)
```

13. Revise que su modelo esté completo

```
modelo.check_model()
```

14. Importe la clase VariableElimination del paquete de inferencia y cree un objeto de esta clase

```
from pgmpy.inference import VariableElimination

infer = VariableElimination(modelo)
```

15. Suponga que Ud selecciona la puerta 1 y el animador la puerta 3. ¿Cuál es la probabilidad de que el carro esté detrás de cada una de las puertas?

```
posterior_p = infer.query(["C"], evidence={"U": 0, "A": 2})
print(posterior_p)
```

Note que los valores se indexan desde 0, luego la puerta 1 corresponde al elemento 0, la 2 al elemento 1, y la 3 al elemento 2.

16. Explique el anterior comando a partir de la documentación del la librería `https://pgmpy.org/exact_infer/ve.html#pgmpy.inference.ExactInference.VariableElimination.query`.

17. Interprete los resultados. ¿Es mejor cambiarse de puerta o no? **R.**

18. Modifique la evidencia para considerar otros casos de selección de puertas, tanto la que Ud selecciona como la que el Animador abre. ¿Cómo interpreta estos resultados? **R**.

19. Compare ahora con el caso en que solo tiene como evidencia la puerta que Ud seleccionó. **R**.

20. Compare ahora con el caso en que solo tiene como evidencia la puerta que abre el Animador. **R**.

21. Concluya. **R**.

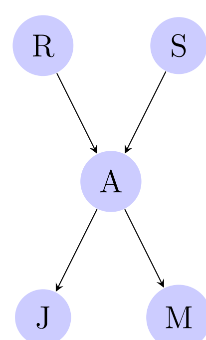
3. El problema de la alarma antirrobo

Ud ha instalado una alarma antirrobo. La alarma es buena detectando intrusos pero también se activa con movimientos sísmicos. Ud tiene además buena relación con dos vecinos, Juan y María, quienes lo llaman en caso de que escuchen la alarma. Juan suele llamar al escuchar la alarma, pero también se confunde con otros ruidos y llama cuando la alarma no ha sonado. María a veces no escucha la alarma. Su evidencia consiste en si Juan y/o María lo llaman, y basado en esto Ud quiere estimar la probabilidad de que haya habido un robo.

1. Para estimar la probabilidad de robo Ud ha decidido modelar este problema como una Red Bayesiana.

2. Defina cinco variables aleatorias R, S, A, J y M que representan el robo, el sismo, la alarma, la llamada de Juan y la de María, respectivamente. Todas las variables tienen espacio de estados $\{V, F\}$ (verdadero y falso).

3. Ud ha definido la siguiente red bayesiana para modelar este problema. En su **reporte** explique por qué decide emplear esta red.



4. Responda las siguientes preguntas en su **reporte**:

    a) ¿Tiene esta red estructuras en V? Si las tiene, ¿Qué implicaciones tienen en la influencia entre las variables de la red?

b) ¿Qué tipo de análisis considera es más relevante en este caso? ¿Causal, evidencial, intercausal? ¿Por qué?

c) ¿Qué independencias condicionales captura esta red?

5. La probabilidad de un robo se ha estimado en 0.01, mientras la de un sismo en 0.02. Modele este comportamiento con las distribuciones de probabilidad $P(R)$ y $P(S)$. **R**.

6. La alarma responde a un robo o a un sismo de acuerdo con la siguiente CPD:

	R	V		F	
A	S	V	F	V	F
V		0.95	0.94	0.29	0.001
F		0.05	0.06	0.71	0.999

   Interprete estas probabilidades en su **reporte**.

7. Si suena la alarma, Juan llama 9 de cada 10 veces, mientras María llama 7 de cada 10 veces. Si la alarma NO suena, Juan llama 1 de cada 20 veces, mientras María llama 1 de cada 100 veces. Modele este comportamiento con las distribuciones de probabilidad $P(J|A)$ y $P(M|A)$. **R**.

8. En su editor de python cree un nuevo archivo alarma.py e implemente el modelo creado para el problema de la alarma antirrobo. Utilice las clases BayesianNetwork, TabularCPD y VariableElimination de *pgmpy*.

9. Una vez tenga el modelo construido, obtenga las independencias con el método

```
print(modelo.get_independencies())
```

   Incluya el resultado en su **reporte** y compare con las que había identificado previamente.

10. Considere 4 casos de evidencia: recibe llamadas tanto de Juan como de María, recibe la llamada de uno solo de ellos, o ninguno lo llama. Calcule la probabilidad de que haya ocurrido un robo en cada caso. Incluya sus resultados y análisis en su **reporte**.

11. Suponga ahora que, cuando NO se activa la alarma, Juan lo llama con la misma probabilidad de María, es decir, 1 de cada 100 veces. Recalcule sus probabilidades y analice los cambios en su **reporte**.