

Universidad Rafael Landívar
Inteligencia Artificial
Primer Semestre 2025

Laboratorio No. 5

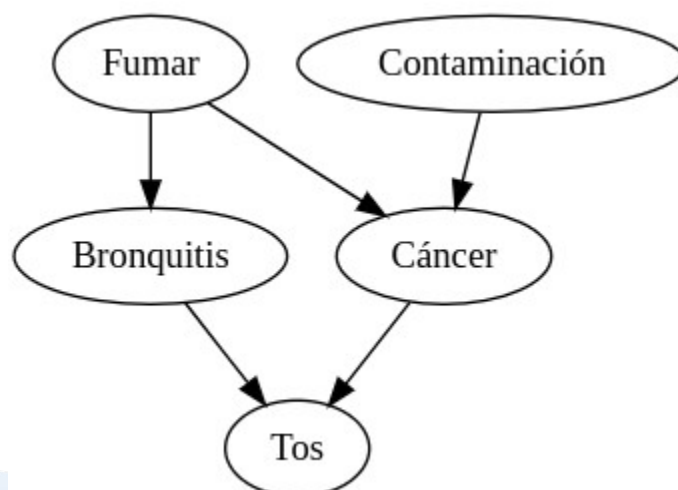
Redes Bayesianas

<https://classroom.github.com/a/x8fRBkbe>

Hace unos años, un equipo de investigadores en salud pública se percató de un alarmante incremento en los casos de enfermedades respiratorias en una gran ciudad. La población no solo presentaba altos índices de tabaquismo, sino que además, la calidad del aire se veía comprometida por la contaminación industrial y del tráfico. Ante esta situación, los expertos decidieron profundizar en las relaciones entre estos factores de riesgo y la aparición de síntomas respiratorios, como la tos.

Para ello, se diseñó un estudio en el que se modelaban de forma probabilística las conexiones entre el hábito de fumar, la exposición a la contaminación, el desarrollo de bronquitis y cáncer de pulmón, y la manifestación de la tos.

La idea era construir un modelo que permitiera predecir la probabilidad de presentar tos basándose en la combinación de estos factores. Este enfoque no solo ayudaría a identificar los principales determinantes de las enfermedades respiratorias, sino que también podría servir para orientar intervenciones tempranas y campañas de prevención, mejorando la calidad de vida de la población.



Fumar (F)	P(Fumar)
+f	0.3
-f	0.7

Contaminacion (C)	P(Contaminacion)
+c	0.4
-c	0.6

Fumar (F)	Bronquitis (B)	P(Bronquitis Fumar)
+f	+b	0.6
+f	-b	0.4
-f	+b	0.3
-f	-b	0.7

Fumar (F)	Contaminacion (C)	Cancer (R)	P(Cáncer Fumar, Contaminación)
+f	+c	+r	0.04
+f	+c	-r	0.96
+f	-c	+r	0.03
+f	-c	-r	0.97
-f	+c	+r	0.02
-f	+c	-r	0.98
-f	-c	+r	0.01
-f	-c	-r	0.99

Bronquitis (B)	Cancer (R)	Tos (T)	P(Tos Bronquitis, Cáncer)
+b	+r	+t	0.9
+b	+r	-t	0.1
+b	-r	+t	0.8
+b	-r	-t	0.2
-b	+r	+t	0.7
-b	+r	-t	0.3
-b	-r	+t	0.1
-b	-r	-t	0.9

Para llevar a cabo el laboratorio deberá utilizar la librería **PyAgrum** (<https://pyagrum.readthedocs.io/en/1.17.2>), puede instalarlo mediante ***pip install***, también se recomienda emplear librerías de visualización en Python, como **matplotlib** o **seaborn**.

Utilizando pyAgrum:

1. Construya la red bayesiana con los 5 nodos descritos, defina su estructura (arcos) y asocie las distribuciones de probabilidad condicional (CPD por sus siglas en inglés) presentadas.
2. Calcular la probabilidad marginal de que una persona presente tos ($P(Tos)$) en la red definida
3. Calcule la probabilidad de que una persona tenga cáncer dado que se sabe que es fumadora
4. Determine la probabilidad de que una persona tenga bronquitis dado que presenta tos
5. Establezca la evidencia de que la persona está expuesta a alta contaminación ($Contaminacion=True$) y calcule:
 1. $P(Cancer=True \mid Contaminacion=True)$
 2. $P(Tos=True \mid Contaminacion=True)$
6. Genere un gráfico comparativo de las probabilidades marginales de Tos en tres escenarios:
 1. Sin evidencia.
 2. Con evidencia de que la persona fuma ($Fumar=True$).
 3. Con evidencia de alta contaminación ($Contaminacion=True$).

Utilice barras o líneas para visualizar las diferencias.

7. Utilizando un muestreador de la red bayesiana (utilizar sampler de PyAgrum), genera un conjunto de datos (por ejemplo, 1000 muestras) y construya un histograma para visualizar la distribución de la variable Tos.
8. Generar un conjunto de datos a partir de la red bayesiana y utiliza un gráfico de dispersión (scatter plot) con jitter (seaborn) para visualizar la relación entre dos variables categóricas, **Fumar** y **Cáncer**.

La entrega es en el repositorio de Github Classroom, y se tratan los siguientes archivos:

- `tarea.py`: Archivo con la implementación de las funciones.
- `enunciado.ipynb`: Notebook con el desarrollo, visualizaciones y explicaciones.
- `Resultados.pdf`: PDF del Notebook solucionado (Agregar al repositorio).

