Estadística aplicada

Probabilidad

Clase 03

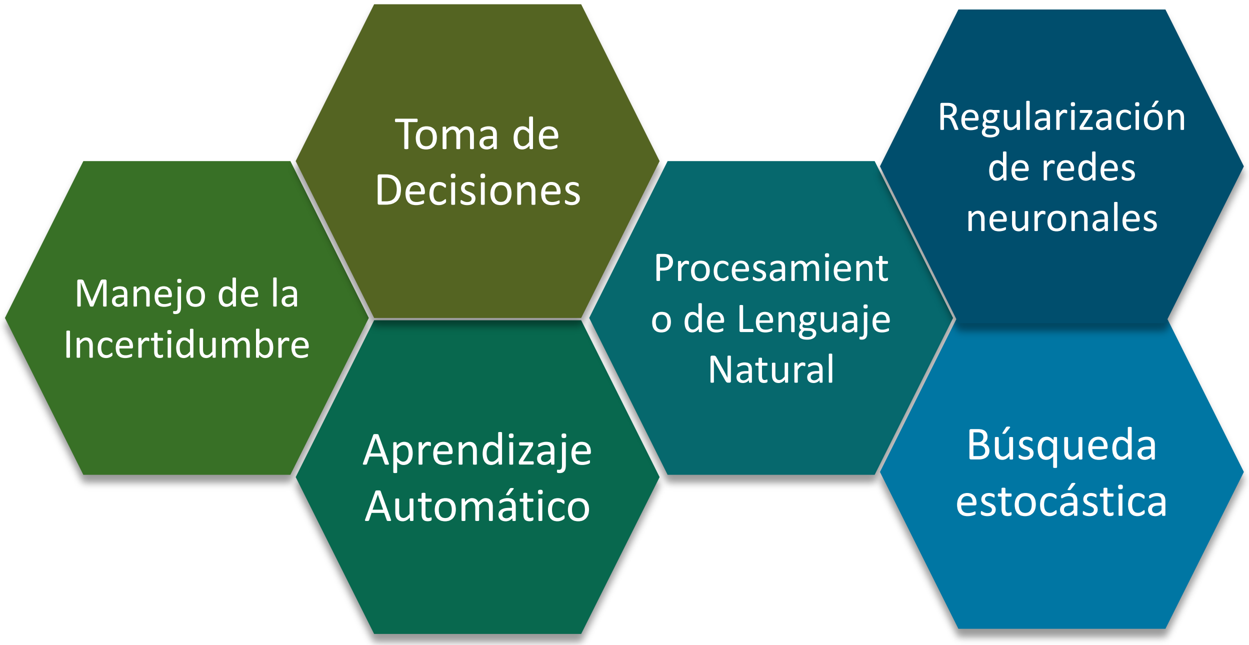
CONQUER BLOCKS

Importancia de la probabilidad en Redes Neuronales

¿Por qué es importante?

La probabilidad es fundamental en la inteligencia artificial (IA) por varias razones, ya que proporciona una base matemática para manejar **la incertidumbre**, tomar decisiones informadas y **aprender de los datos**

Aplicaciones de la probabilidad en IA

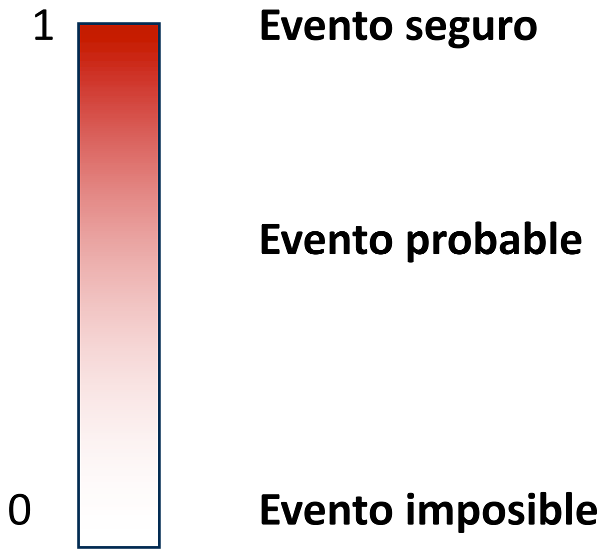


¿Qué es la probabilidad?

¿Qué es la probabilidad?

La probabilidad es una rama de las matemáticas que estudia la incertidumbre y la aleatoriedad. La teoría de la probabilidad proporciona un marco formal para cuantificar y analizar eventos inciertos, permitiendo predecir la ocurrencia de eventos futuros y entender la variabilidad inherente en los procesos.

¿Cómo se mide la probabilidad



Conceptos básicos

Experimento Aleatorio

Un proceso que lleva a un resultado incierto. Por ejemplo, lanzar una moneda, tirar un dado, o seleccionar una carta de una baraja.

Espacio Muestral (S)

El conjunto de todos los resultados posibles de un experimento aleatorio. Por ejemplo, para el lanzamiento de una moneda, el espacio muestral es {Cara, Cruz}.

Evento

Un subconjunto del espacio muestral. Es el resultado o un conjunto de resultados que nos interesa. Por ejemplo, obtener una cara al lanzar una moneda.

Probabilidad de un evento P(E)

La medida que cuantifica la posibilidad de que ocurra el evento. Para un evento E, la probabilidad P(E) se calcula como:

$$P(E) = \frac{\text{Número de resultados favorables}}{\text{Número total de resultados posibles}}$$

Axiomas

¿Qué es un Axioma?

Un axioma es una proposición o declaración que se acepta como verdadera sin necesidad de demostración dentro de un sistema lógico o matemático. Los axiomas sirven como puntos de partida para deducir y construir otros teoremas y resultados en dicho sistema.

¿Características de un Axioma?

1. **Autoevidencia**: Aunque no siempre son obvios, los axiomas son considerados como declaraciones básicas que no requieren prueba.
2. **Fundamentales**: Actúan como las bases sobre las cuales se construye una teoría o sistema.
3. **Consistentes**: En un sistema bien definido, los axiomas no deben contradecirse entre sí.

Ejemplos

1. **"A través de dos puntos distintos pasa una única recta". –** Geometría Euclídea
2. **"Si A es verdadera, entonces A es verdadera" --** Principio de identidad.

Axiomas en la probabilidad

¿Qué es un Axioma?

Los axiomas en probabilidad son afirmaciones básicas que sabemos que son ciertas por definición

3 axiomas de la probabilidad

No Negatividad

La probabilidad de un evento es un número no negativo.

$\geq 0 P(E) \geq 0$

Normalización

La probabilidad del espacio muestral completo es 1.

$P(S)=1$

Adición

Si dos eventos $E1$ y $E2$ son mutuamente excluyentes (no pueden ocurrir al mismo tiempo), la probabilidad de que ocurra $E1$ o $E2$ es la suma de sus probabilidades individuales.

$P(E1 \cup E2)=P(E1)+P(E2)$

Teoremas básicos: Regla de la Suma

Regla de la suma

Describe cómo calcular la probabilidad de la **unión de dos o más eventos**. Esta regla tiene dos formas principales: una para **eventos mutuamente excluyentes** (eventos que no pueden ocurrir simultáneamente) y otra para eventos **no mutuamente excluyentes** (eventos que pueden ocurrir simultáneamente).

Eventos mutuamente excluyentes

Si dos eventos A y B son mutuamente excluyentes, significa que no pueden ocurrir al mismo tiempo. En este caso, la probabilidad de que ocurra A o B es simplemente la suma de sus probabilidades individuales.

$P(A \cup B) = P(A) + P(B)$

Si Lanzamos un dado de seis caras. Queremos encontrar la probabilidad de obtener un 2 o un 5. Los eventos "obtener un 2" y "obtener un 5" **son mutuamente excluyentes**, porque no se pueden obtener ambos resultados al mismo tiempo.

$$P(2) = \tfrac{1}{6}$$
$$P(5) = \tfrac{1}{6}$$
$$P(2 \text{ o } 5) = P(2) + P(5) = \tfrac{1}{6} + \tfrac{1}{6} = \tfrac{2}{6} = \tfrac{1}{3}$$

Teoremas básicos: Regla de la Suma

Eventos NO mutuamente excluyentes

Si dos eventos *A* y *B* **no** son mutuamente excluyentes, significa que **pueden ocurrir al mismo tiempo.** En este caso, la probabilidad de que ocurra *A* o *B* es la suma de sus probabilidades individuales menos la probabilidad de que ocurran ambos eventos al mismo tiempo (su intersección).

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Supongamos que tenemos una baraja estándar de 52 cartas. Queremos encontrar la probabilidad de sacar una carta que sea un rey o un corazón. Los eventos "sacar un rey" y "sacar un corazón" no son mutuamente excluyentes porque hay un rey de corazones que pertenece a ambos eventos.

$$P(\text{rey}) = \frac{4}{52} = \frac{1}{13}$$
$$P(\text{corazón}) = \frac{13}{52} = \frac{1}{4}$$
$$P(\text{rey y corazón}) = P(\text{rey de corazones}) = \frac{1}{52}$$
$$P(\text{rey o corazón}) = P(\text{rey}) + P(\text{corazón}) - P(\text{rey y corazón})$$
$$P(\text{rey o corazón}) = \frac{1}{13} + \frac{1}{4} - \frac{1}{52} = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$$

Teoremas básicos: Regla del PRODUCTO

Regla del producto

describe cómo calcular la probabilidad de la **intersección de dos o más eventos**. Esta regla tiene dos formas principales: una para **eventos independientes** y otra para **eventos dependientes**

Regla del Producto para Eventos Independientes

Si dos eventos A y B son independientes, significa que **la ocurrencia de uno no afecta la probabilidad de ocurrencia del otro**. En este caso, la probabilidad de que ambos eventos ocurran es **el producto de sus probabilidades individuales**.

$P(A \cap B) = P(A) \times P(B)$

Supongamos que lanzamos una moneda y un dado. Queremos encontrar la probabilidad de obtener una cara en la moneda y un 3 en el dado. Estos eventos son independientes porque el resultado del lanzamiento de la moneda no afecta el resultado del lanzamiento del dado.

$$P(\text{cara}) = \frac{1}{2}$$
$$P(3) = \frac{1}{6}$$
$$P(\text{cara y } 3) = P(\text{cara}) \times P(3) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$$

Teoremas básicos: Regla del PRODUCTOS

Regla del Producto para Eventos Dependientes

Si dos eventos A y B son dependientes, significa que la ocurrencia de uno afecta la probabilidad de ocurrencia del otro. En este caso, **la probabilidad de que ambos eventos ocurran es el producto de la probabilidad de uno de los eventos y la probabilidad condicional del otro evento dado que el primero ha ocurrido**.

$P(A \cap B) = P(A) \times P(B/A)$

Donde:
$P(B/A)$ es la probabilidad condicional de B dado que A ha ocurrido.

Supongamos que tenemos una baraja inglesa de 52 cartas. Queremos encontrar la probabilidad de sacar un as y luego, sin reemplazar la carta, sacar un rey. Estos eventos son dependientes porque el resultado del primer evento (sacar un as) afecta la probabilidad del segundo evento (sacar un rey).

$$P(\text{as}) = \frac{4}{52} = \frac{1}{13}$$

Después de sacar un as, quedan 51 cartas en la baraja, de las cuales 4 son reyes.

$$P(\text{rey} \mid \text{as}) = \frac{4}{51}$$

Entonces, la probabilidad de sacar un as y luego un rey es:

$$P(\text{as y rey}) = P(\text{as}) \times P(\text{rey} \mid \text{as}) = \frac{1}{13} \times \frac{4}{51} = \frac{4}{663}$$

Distribuciones de Probabilidad

¿Qué es una distribución de probabilidad?

Una distribución de probabilidad es una **función** que asigna **a cada evento o resultado posible** de una variable aleatoria una **probabilidad de ocurrencia**. Esta distribución puede ser de dos tipos principales: **discreta o continua**.

Distribuciones discretas

Las distribuciones discretas se aplican a variables aleatorias que pueden tomar un número finito o contable de valores

- **Distribución Binomial**
- **Distribución Poisson**

Distribuciones continuas

Las distribuciones continuas se aplican a variables aleatorias que pueden tomar **cualquier valor** dentro de un rango continuo

- **Distribución Normal**
- **Distribución Uniforme**

Distribuciones de Probabilidad

Variables de las distribuciones

- **Esperanza Matemática (Media)**: Es el valor promedio esperado de la variable aleatoria y se denota como $E(X)$. Para una variable **discreta**, es la **suma de los productos de cada valor por su probabilidad**. Para una variable **continua**, **es la integral del producto del valor por su densidad de probabilidad**.

- **Varianza**: Mide la **dispersión** de los valores alrededor de la media. Se denota como $Var(X)$ y es el valor esperado del cuadrado de la desviación de la variable respecto a su media.

- **Desviación Estándar**: Es la raíz cuadrada de la varianza y proporciona una medida de dispersión en las mismas unidades que la variable original.

Distribuciones Discretas: Poisson

Distribuciones de Poisson

Modela el número de eventos que pasan en un intervalo fijo de tiempo o espacio, bajo estas condiciones:

Los eventos son independientes: La ocurrencia de un evento no afecta la probabilidad de que ocurra otro evento.
La tasa de ocurrencia de eventos es constante: No cambia durante el intervalo considerado.
Los eventos ocurren uno a uno: No pueden ocurrir simultáneamente

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

- X es la variable aleatoria que representa el número de eventos.
- k es el número específico de eventos.
- λ es el promedio esperado de eventos en el intervalo (también llamado parámetro de tasa).
- e es la base del logaritmo natural (aproximadamente 2.71828).

E(X)=λ
Var(X)= λ

5!=5·4·3·2·1 → 5!=120

Ejemplo

Supongamos que un centro de llamadas recibe en promedio 5 llamadas por minuto (λ=5). Queremos encontrar la probabilidad de que reciban exactamente 3 llamadas en un minuto.

$$P(X = 3) = \frac{5^3 e^{-5}}{3!} = 0{,}1404 \rightarrow 14\%$$

Distribuciones Discretas: Poisson

Ejercicio

En una red neuronal, el *dropout* es una técnica que se utiliza para prevenir el sobreajuste. Consiste en desactivar aleatoriamente una fracción de neuronas durante el entrenamiento. Supongamos que la probabilidad de que una neurona sea desactivada sigue una distribución de Poisson con una tasa de $\lambda=0.5$

1. Simula el proceso de *dropout* para una capa con 100 neuronas utilizando la distribución de Poisson.
2. Calcula y muestra el número medio de neuronas desactivadas en 1000 iteraciones.
3. Grafica la distribución de neuronas desactivadas en estas 1000 iteraciones.

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Distribucion_poisson.py

Distribuciones Discretas: Binomial

Distribuciones de Binomial

La distribución binomial modela el número de éxitos en una serie de ensayos independientes, donde cada ensayo tiene dos posibles resultados: éxito o fracaso. Funciona para los siguientes supuestos:

Número Fijo de Ensayos: Hay un número fijo n de ensayos.
Independencia: Los ensayos son independientes entre sí.
Probabilidad Constante: La probabilidad de éxito *p* es constante en cada ensayo.
Resultados Binarios: Cada ensayo tiene solo dos resultados posibles: éxito (1) o fracaso (0)

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

E(X)=n·p
Var(X)= n·p·(1−p)

- X es la variable aleatoria que representa el número de éxitos.
- k es el número específico de éxitos.
- n es el número de ensayos.
- p es la probabilidad de éxito en cada ensayo.
- $\binom{n}{k}$ es el coeficiente binomial, que se calcula como: $\binom{n}{k}=\frac{n!}{k!(n-k)!}$

Ejemplo

Supongamos que lanzas una moneda 10 veces n=10 y quieres encontrar la probabilidad de obtener exactamente 6 caras (*k*=6). La probabilidad de obtener una cara en un lanzamiento de moneda justo es *p*=0.5.

Usamos la fórmula de la PMF de Binomial:

$$P(X=6)=\binom{10}{6}(0.5)^6(0.5)^4=\binom{10}{6}(0.5)^{10}$$
$$\binom{10}{6}=\frac{10!}{6!4!}=210$$
$$P(X=6)=210\cdot(0.5)^{10}=210\cdot\frac{1}{1024}\approx 0.205$$

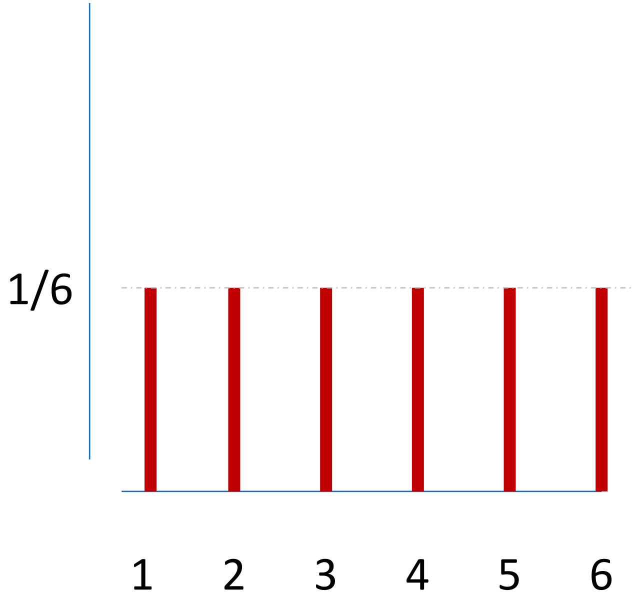
Distribuciones de Probabilidad: Continuas

¿Cual es la diferencia entre una distribución discreta y una continua?

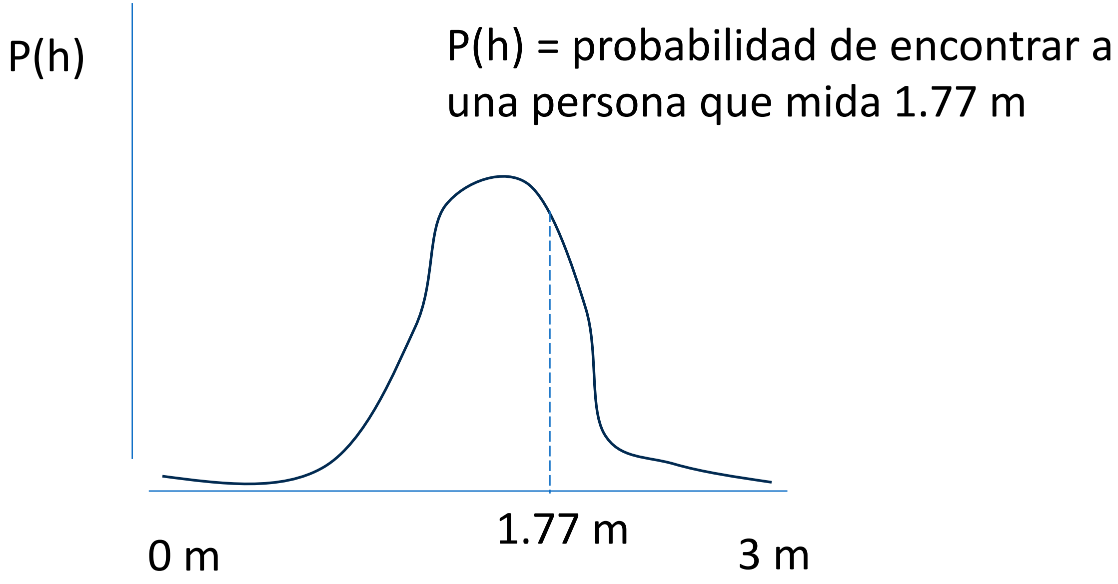
En la continua, la probabilidad esta definida para infinitos eventos, sin embargo para las discretas solo están definidos para unos pocos

Discreta

Tirar un dado



Altura de la población mundial



No se encuentra definido la probabilidad de que salga en el dado 1,53

La función se encuentra definida para todos los valores de su dominio

Distribuciones de Probabilidad: Continuas

Variables de las distribuciones

•**Media (μ):** Para una variable **discreta**, es la **suma de los productos de cada valor por su probabilidad**. Para una variable **continua, es la integral del producto del valor por su densidad de probabilidad**.

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

•**Varianza**: Mide la **dispersión** de los valores alrededor de la media. Se denota como *Var*(*X*) y es el valor esperado del cuadrado de la desviación de la variable respecto a su media.

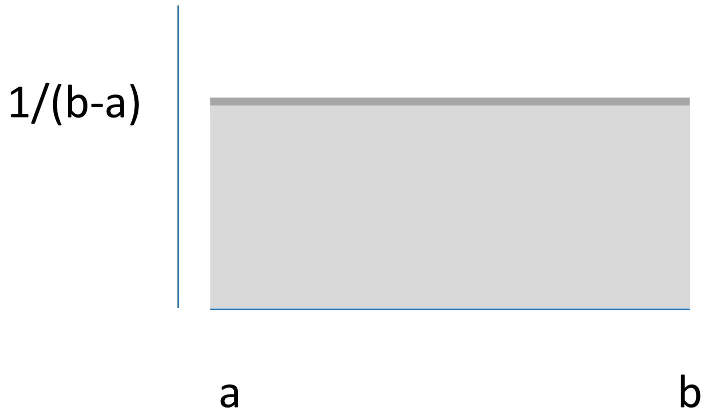
•**Desviación Estándar**: Es la raíz cuadrada de la varianza y proporciona una medida de dispersión en las mismas unidades que la variable original.

Distribuciones de Probabilidad: Uniforme

Distribuciones de probabilidad Uniforme

La distribución uniforme continua se utiliza para modelar situaciones en las que una variable aleatoria puede tomar cualquier valor dentro de un intervalo [a,b], y cada uno de estos valores es igualmente probable.

$$P(a \leq X \leq b) = \frac{1}{b-a}$$



La probabilidad es proporcional a la longitud del intervalo [a, b].

Ejemplo

Giro una ruleta con una fuerza aleatoria. ¿Cuál es la probabilidad de que la ruleta pare en ángulo=30

$$P(X = 30) = \frac{30-30}{360-0} = 0$$

¿Y entre el ángulo=20 y 30?

$$P(X = 30) = \frac{30-20}{360-0} = 3{,}6\%$$

Distribuciones de Probabilidad: Normal

Distribución Normal

También conocida como distribución de **Gauss**, es una de las distribuciones de **probabilidad más importantes** en la estadística. La gráfica de la distribución normal es una curva en forma de campana simétrica alrededor de la media. Esto significa que el 50% de los valores se encuentran a la izquierda de la media y el 50% a la derecha

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- $f(x)$ es la densidad de probabilidad en el valor x.
- μ es la media de la distribución.
- σ es la desviación estándar.
- $\exp$ es la función exponencial.
- π es el número pi (aproximadamente 3.14159).

En una distribución normal de media 0 y desviación estándar 1, se puede saber que los valores van a estar distribuidos de la siguiente manera

- 68% de los datos caen dentro de 1 desviación estándar de la media ($\mu \pm \sigma$)
- 95% de los datos caen dentro de 2 desviaciones estándar de la media ($\mu \pm 2\sigma$).
- 99.7% de los datos caen dentro de 3 desviaciones estándar de la media ($\mu \pm 3\sigma$).

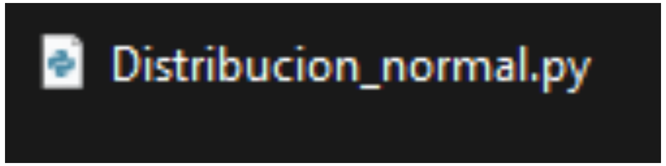
Distribuciones de Probabilidad: Normal

Ejemplo en python

Las alturas de una clase de estudiantes se puede explicar a través de una distribución normal con media 170cm y desviación estándar 12cm

Calcula la probabilidad de que un alumno mida entre 189 y 190 cm

Calcula también la probabilidad de que un alumno mida entre 174 y 175cm



Para que se usa en inteligencia artificial

La distribución normal es una herramienta fundamental en inteligencia artificial debido a su capacidad para modelar datos continuos, estimar parámetros, manejar incertidumbre y formar la base de muchos algoritmos y técnicas de análisis de datos y toma de decisiones.

Inicio de parámetros, filtros de señales, Redes bayesianas etc