

# Modelos Estocásticos para Manufactura y Servicios (INDG-1008): **Unidad 02**

Luis I. Reyes Castro

Escuela Superior Politécnica del Litoral (ESPOL)  
Guayaquil - Ecuador

2017 - Primer Término

- 1 Árboles de Decisión
- 2 Inferencia Bayesiana (Experimentación)
- 3 Árboles de Decisión con Experimentación

- 1 Árboles de Decisión
- 2 Inferencia Bayesiana (Experimentación)
- 3 Árboles de Decisión con Experimentación

# Árboles de Decisión

- 1 Árboles de Decisión
- 2 Inferencia Bayesiana (Experimentación)
- 3 Árboles de Decisión con Experimentación

# Inferencia Bayesiana (Experimentación)

Como motivación, empecemos considerando la **Paradoja del Test**.

- Un raro tipo de cáncer ha sido descubierto hace poco. Se estima que el 0.00001% de la población es genéticamente vulnerable a este cáncer.
- Un test para comprobar la vulnerabilidad genética a este nuevo tipo de cáncer acaba de ser desarrollado. El desempeño del test es como sigue:
  - Si el paciente es genéticamente vulnerable, el test arroja resultado positivo el 98% de las veces.
  - Si el paciente no es vulnerable, el test arroja resultado negativo el 94% de las veces.
- Usted se acaba de hacer el test, y el resultado fue positivo. Cuál es la probabilidad de que usted realmente sea genéticamente vulnerable?

## Problema Prototipo:

- Dos variables aleatorias:
  - Una variable aleatoria  $X$  que representa la categoría real pero desconocida. Usualmente decimos que  $X$  es la variable no observada.
  - Una variable aleatoria  $Y$  que representa el resultado de un test o clasificador que depende probabilísticamente en  $X$ . Usualmente decimos que  $Y$  es la variable observada.
- Se nos provee la distribución de la variable no observada  $X$ . *I.e.:*

$$\forall x \in \text{soporte}(X) : \mathbb{P}(X = x)$$

- Usualmente se la denomina *distribución a priori*.

# Inferencia Bayesiana (Experimentación)

- También se nos provee la distribución de la variable observada  $Y$  condicional en la variable no observada  $X$ . *I.e.*:

$$\forall x \in \text{soporte}(X), \forall y \in \text{soporte}(Y) : \mathbb{P}(Y = y \mid X = x)$$

- En el contexto de clasificadores, usualmente conocemos a esta distribución como la *matriz de confusión*.
- Nuestro objetivo es calcular la distribución de la variable no observada  $X$  condicional en la variable observada  $Y$ . *I.e.*,

$$\forall x \in \text{soporte}(X), \forall y \in \text{soporte}(Y) : \mathbb{P}(X = x \mid Y = y)$$

- Usualmente se la denomina *distribución posterior predictiva*.



- Para lograr nuestro objetivo generalmente necesitamos calcular con anterioridad la distribución marginal de la variable observada  $Y$ . *I.e.*,

$$\forall y \in \text{soporte}(Y) : \mathbb{P}(Y = y)$$

- Usualmente se la denomina *distribución posterior*.
- Una vez calculada...

- 1 Árboles de Decisión
- 2 Inferencia Bayesiana (Experimentación)
- 3 Árboles de Decisión con Experimentación

# Árboles de Decisión con Experimentación