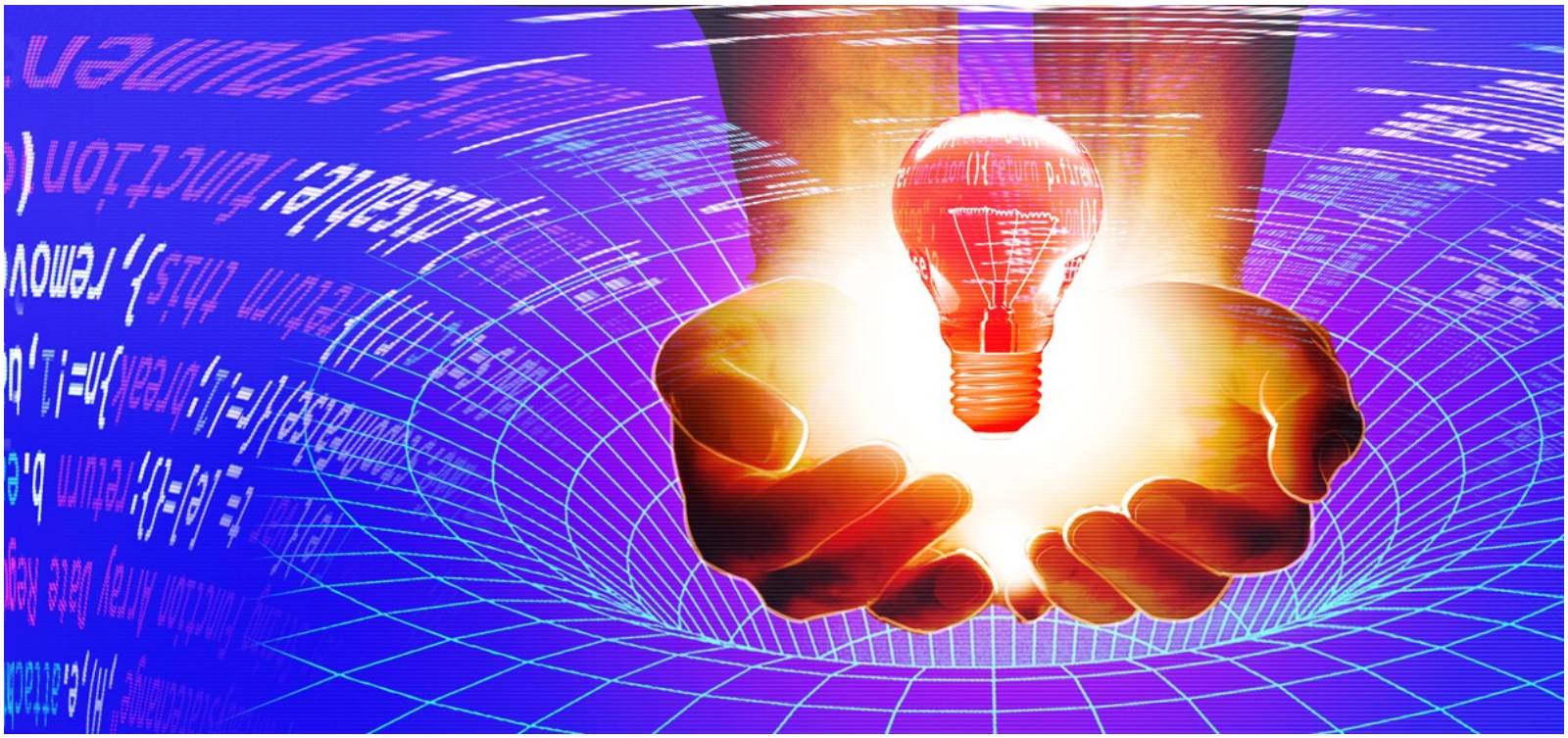


Ciencia de datos energéticos

Ing. Oscar Germán Duarte Velasco M.Sc., Ph.D.

Universidad Nacional de Colombia
Facultad de Ingeniería
Departamento de Ingeniería Eléctrica y Electrónica
Sede Bogotá

Bogotá, septiembre del 2023

Organiza
Dirección Nacional de Innovación Académica

coursera

UNIVERSIDAD NACIONAL DE COLOMBIA

Derechos de autor

Las lecturas del curso Ciencias de datos energéticos están publicadas bajo una licencia Creative Commons:

Reconocimiento-NoComercial-CompartirIgual 3.0 Unported (CC BY-NC-SA 3.0)

Usted es libre de:
Compartir - copiar y redistribuir el material en cualquier medio o formato. Adaptar - remezclar, transformar y crear a partir del material.

Bajo las condiciones siguientes:

BY	**Reconocimiento —** Debe reconocer adecuadamente la autoría, proporcionar un enlace a la licencia e indicar si se han realizado cambios. Puede hacerlo de cualquier manera razonable, pero no de una manera que sugiera que tiene el apoyo del licenciador o lo recibe por el uso que hace.
NC	**NoComercial —** No puede utilizar el material para una finalidad comercial.
SA	**CompartirIgual —** Si remezcla, transforma o crea a partir del material, deberá difundir sus contribuciones bajo la misma licencia que el original.
No hay restricciones adicionales — No puede aplicar términos legales o medidas tecnológicas que legalmente restrinjan realizar aquello que la licencia permite.	
Texto completo de la licencia en: https://creativecommons.org/licenses/by-nc-sa/3.0/deed.es_ES	

Organiza
Dirección Nacional de Innovación Académica

coursera

UNIVERSIDAD NACIONAL DE COLOMBIA

Créditos curso Ciencia de datos energéticos

PhD. Oscar Germán Duarte Velasco
Profesor titular
Departamento de Ingeniería Eléctrica y Electrónica

M. Sc. Diana Milena Jaramillo Quiceno
Coordinadora convenio Coursera - Universidad Nacional de Colombia Asesora pedagógica, metodológica y de políticas de accesibilidad.

Agradecimientos:

M. Sc. Gabriel Ernesto Barrero Tapias
Director Nacional de Innovación Académica

Juan David Alejandro Becerra Becerra
Editor y Realizador

Daniel Reyes Leguizamón
Diseñador gráfico CE-Lab

Unimedios Medellín
Imagen Institucional - Unimedios Bogotá
Vicerrectoría Académica
Universidad Nacional de
Colombia 2023

Ciencia de datos energéticos

Contenido del módulo 4

Oscar Duarte, Ph.D.

Universidad Nacional de Colombia
Departamento de Ingeniería Eléctrica y Electrónica

Estructura del curso

	Tema de energía.	**Tema de informática.**
1	La cadena de la electricidad.	El proceso de la ciencia de datos.
2	Potencia y energía en CA. Sistemas Trifásicos.	Comprensión de datos.
3	Consumo de Energía. Eficiencia energética.	Preparación de datos.
4	Transporte de Energía Eléctrica.	Descubrimiento de patrones. Clasificación.
5	Generación de Energía Eléctrica.	Descubrimiento de explicaciones. Regresiones.
6	Comercialización y mercados de Energía.	Series de tiempo.

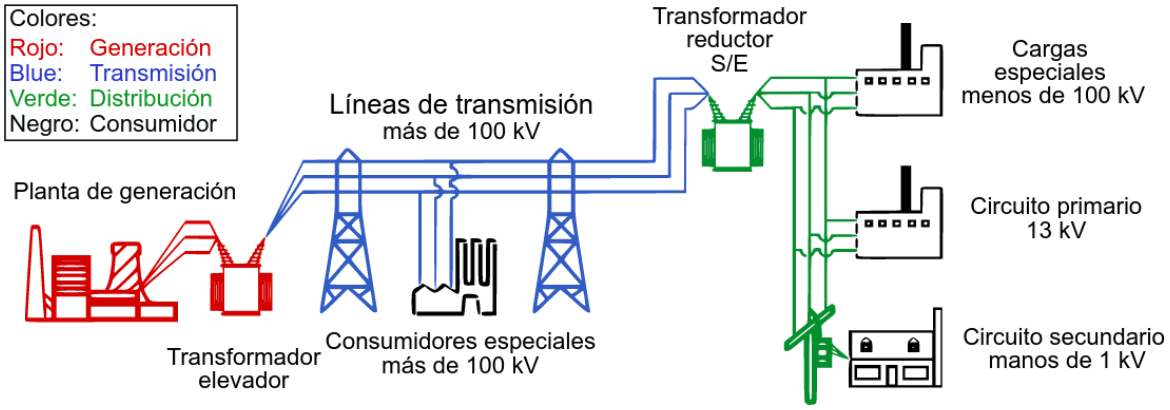
Ciencia de datos energéticos

Transporte de energía eléctrica - 1

Oscar Duarte, Ph.D.

Universidad Nacional de Colombia
Departamento de Ingeniería Eléctrica y Electrónica

La cadena de la electricidad

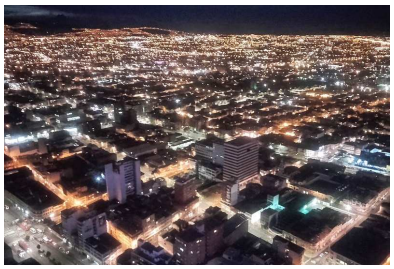


Transporte de energía
Transmisión.

Fuente de energía

Medio de transporte

Uso de energía (carga)



Transporte de energía
Distribución.

Fuente de energía

Medio de transporte

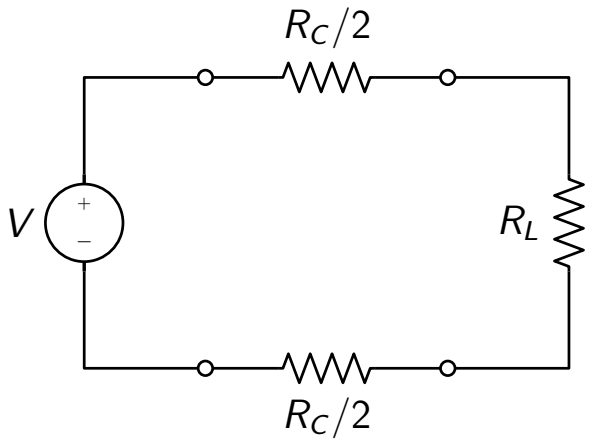
Uso de energía (carga)

Preocupaciones a atender en el transporte de energía

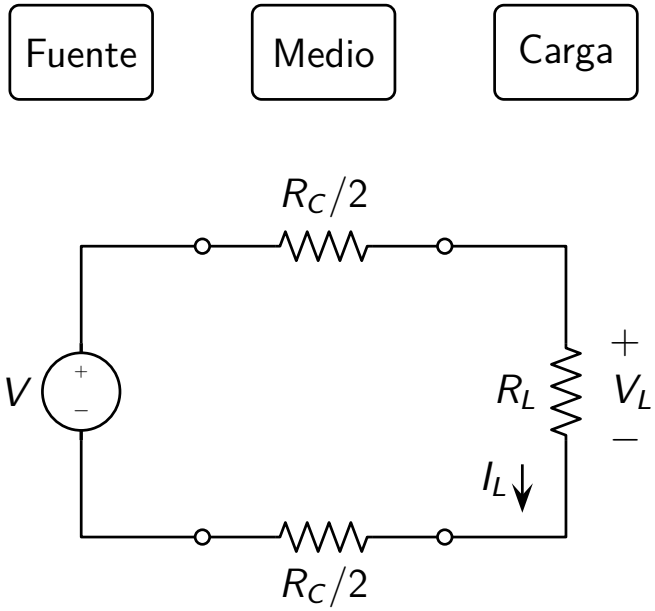
1. Capacidad de transporte.
2. Calidad de potencia.
3. Eficiencia energética.
4. Seguridad:
   - Sobrecargas.
   - Cortocircuitos.
   - Sobretensiones.
5. Atención a fallas.

Capacidad de transporte

Fuente Medio Carga



Capacidad de transporte

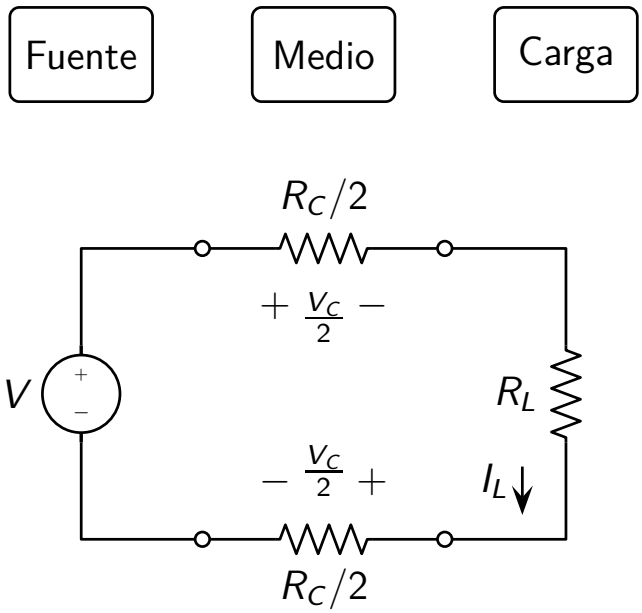


$$I_L = \frac{V}{R_T} = \frac{V}{R_C + R_L}$$

$$V_L = I_L R_L = V \frac{R_L}{R_C + R_L}$$

$$P_L = I_L V_L = V^2 \frac{R_L}{(R_C + R_L)^2}$$

Capacidad de transporte

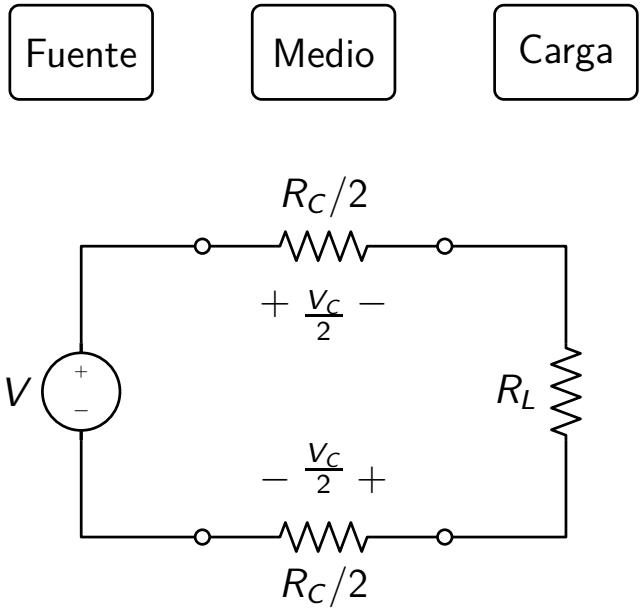


$$I_L = \frac{V}{R_T} = \frac{V}{R_C + R_L}$$

$$V_C = I_L R_C = V \frac{R_C}{R_C + R_L}$$

$$P_C = I_L V_C = V^2 \frac{R_C}{(R_C + R_L)^2}$$

Capacidad de transporte



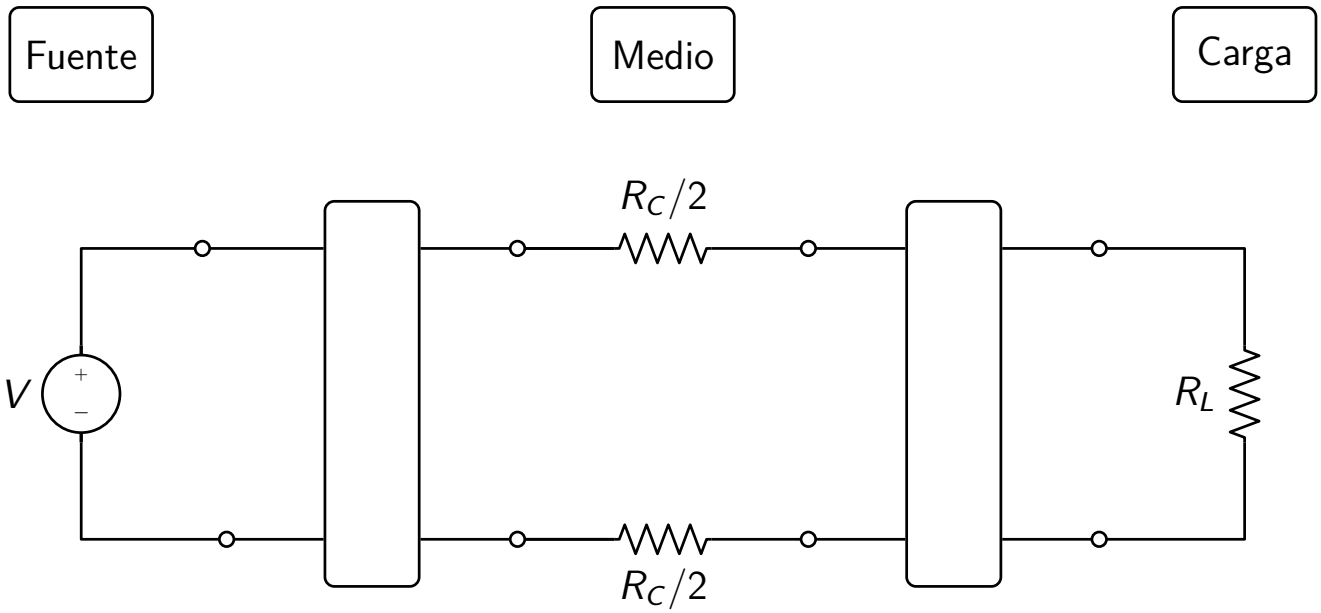
¿Cómo disminuir I_L, R_C?

$$I_L = \frac{V}{R_T} = \frac{V}{R_C + R_L}$$

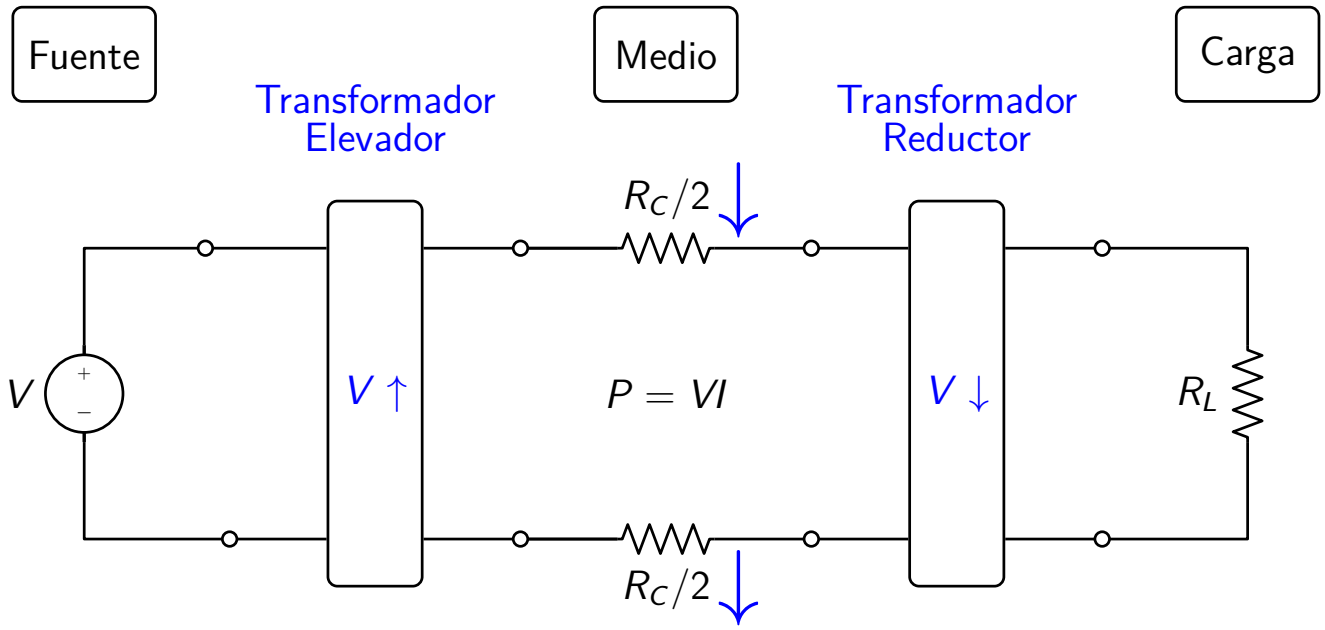
$$V_C = I_L R_C = V \frac{R_C}{R_C + R_L}$$

$$P_C = I_L V_C = V^2 \frac{R_C}{(R_C + R_L)^2}$$

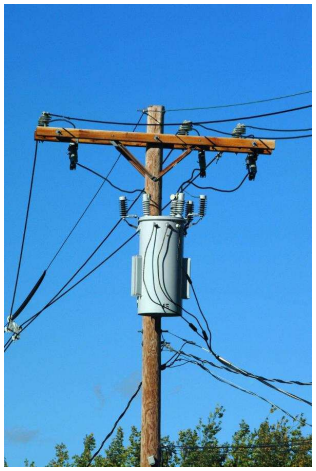
Capacidad de transporte



Capacidad de transporte

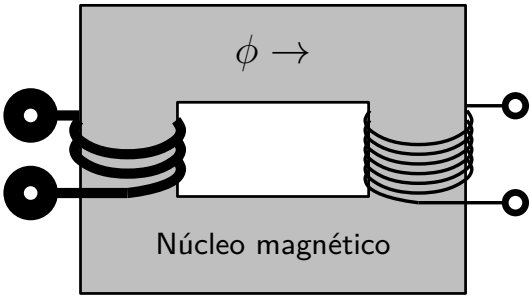


Transformación de energía eléctrica

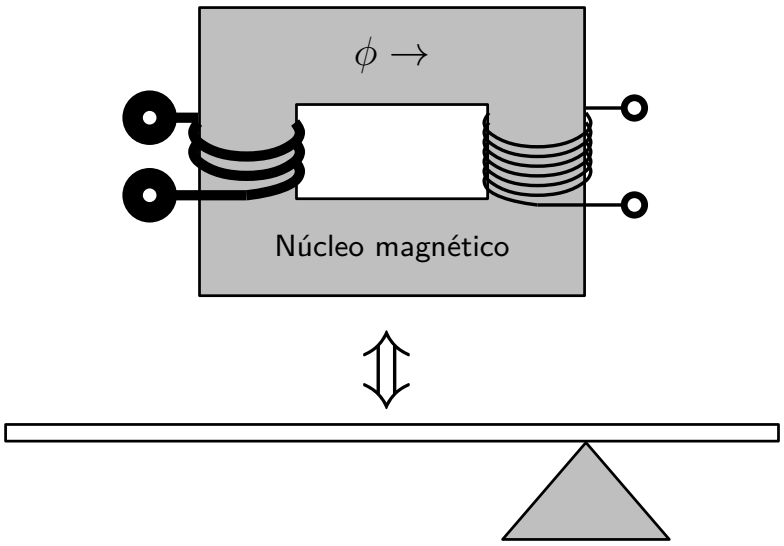


$\Leftrightarrow$

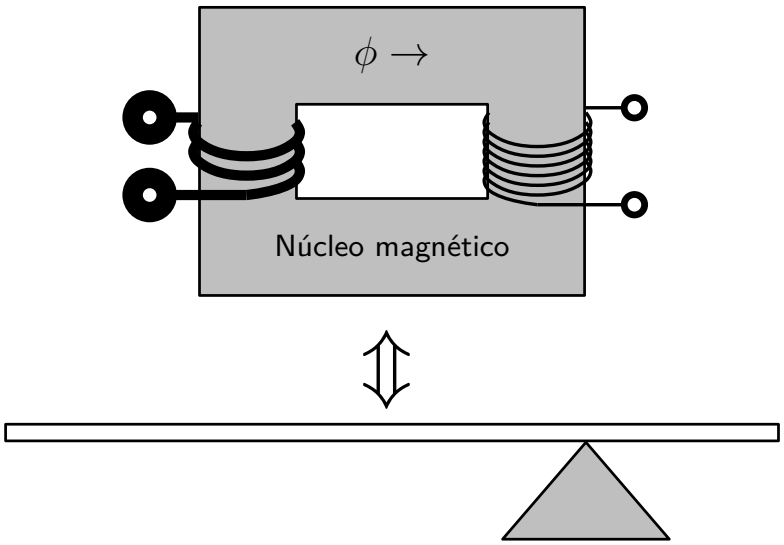
Transformación de energía eléctrica



Transformación de energía eléctrica



Transformación de energía eléctrica



Fuerza

Velocidad

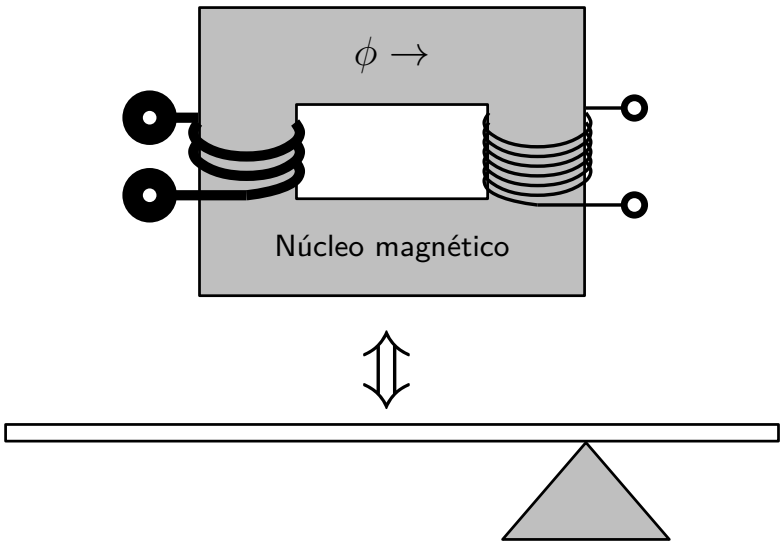
Fuerza

Velocidad

Transformación de energía eléctrica

Tensión

Corriente



Tensión

Corriente

Fuerza

Velocidad

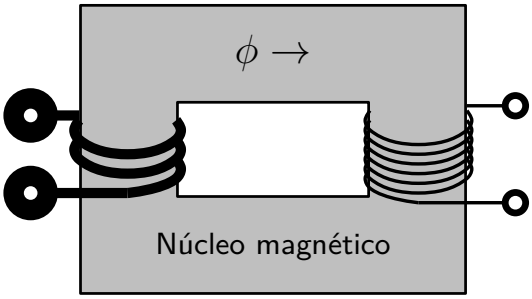
Fuerza

Velocidad

Transformación de energía eléctrica

Tensión

Corriente



Tensión

Corriente

Potencia = Tensión * Corriente

Conservación de potencia

Transformación de energía eléctrica

- ¿Qué es preferible?
  - Tensión baja y corriente alta ó
  - Tensión alta y corriente baja

Transformación de energía eléctrica

- ¿Qué es preferible?
  - Tensión baja y corriente alta ó
  - Tensión alta y corriente baja
- En el consumo:
  - Tensión baja
  - Las altas tensiones son peligrosas
- Para transportar energía:
  - Corriente baja
  - Conductores más delgados

Capacidad de Transporte
Niveles de tensión

	Tensión	Distancias
Generación	$< 10kV$	
Transmisión	$> 200kV$	$> 100km$
Distribución	$< 20kV$	$\approx 10km$
Consumo	$< 0.5kV$	

¡Gracias!

ogduartev@unal.edu.co
@ogduartev

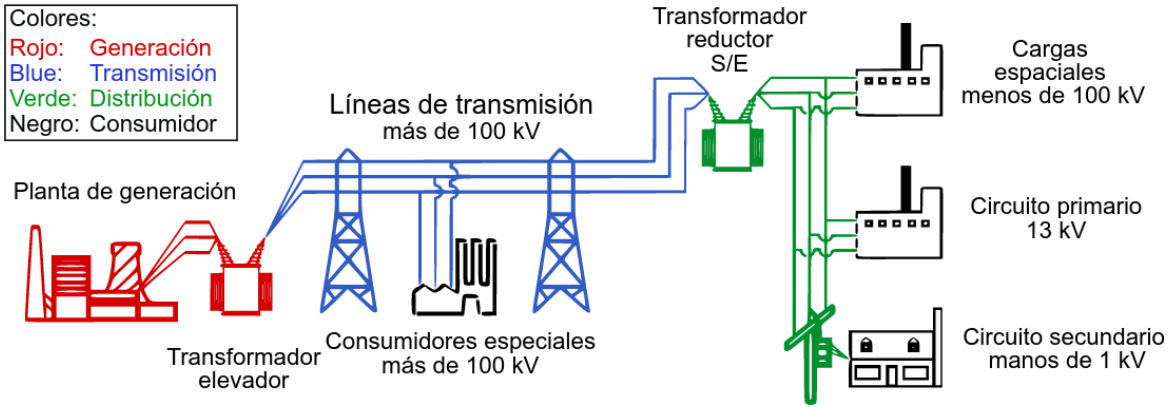
Ciencia de datos energéticos

Transporte de energía eléctrica - 2

Oscar Duarte, Ph.D.

Universidad Nacional de Colombia
Departamento de Ingeniería Eléctrica y Electrónica

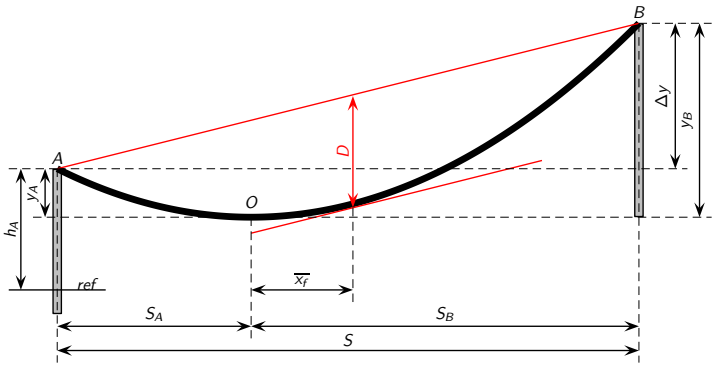
La cadena de la electricidad

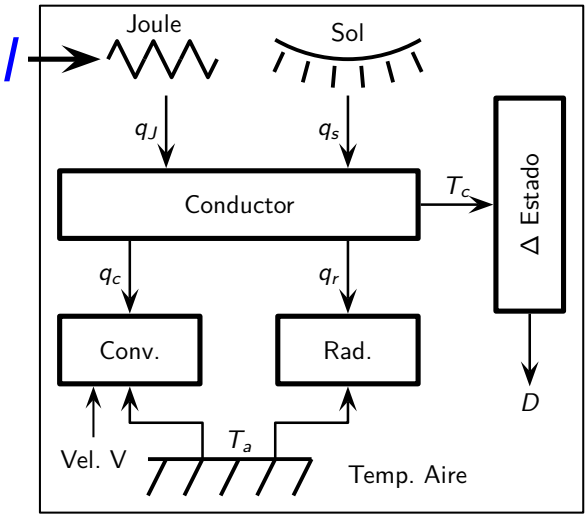


Capacidad de Transporte
Calentamiento

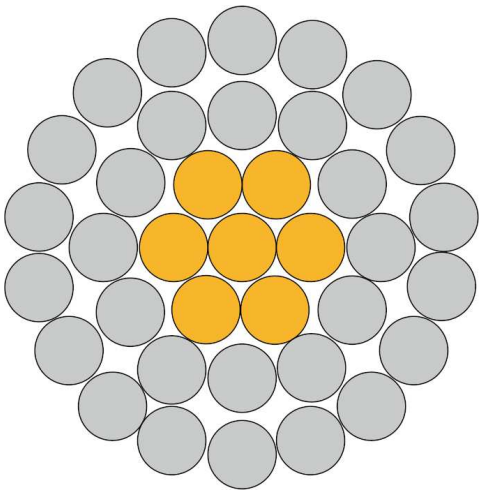
Capacidad de Transporte
Calentamiento

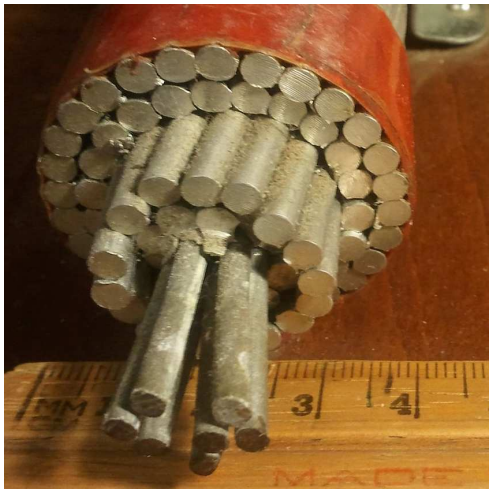
Capacidad de Transporte



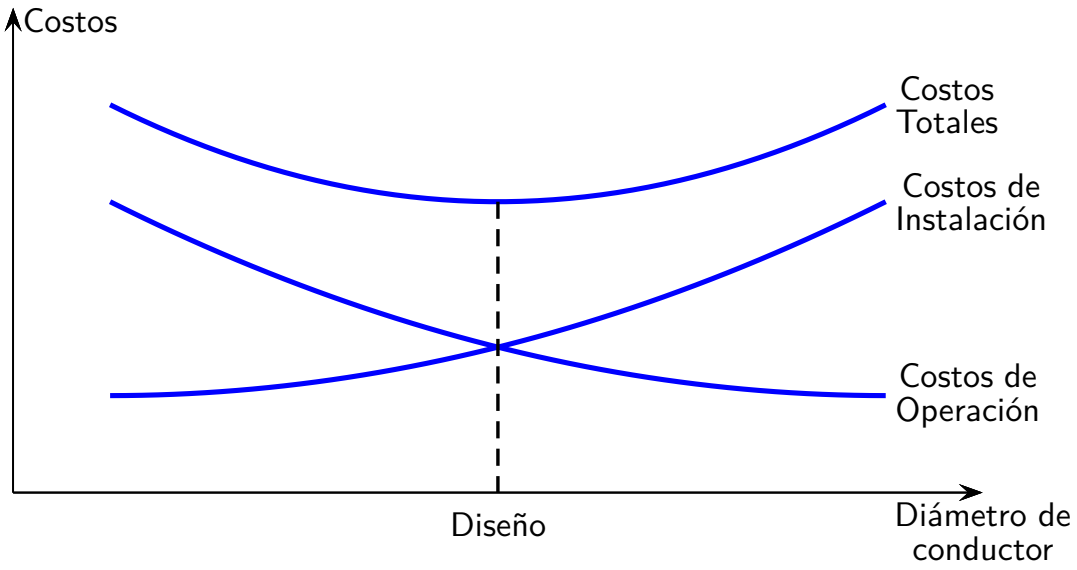


Capacidad de transporte
El conductor

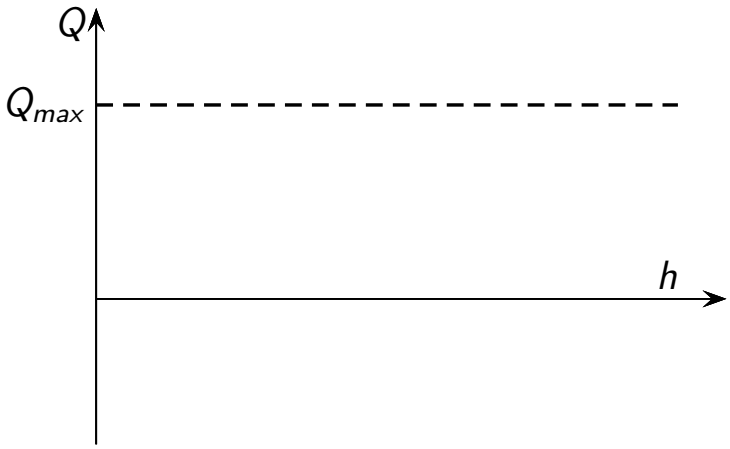




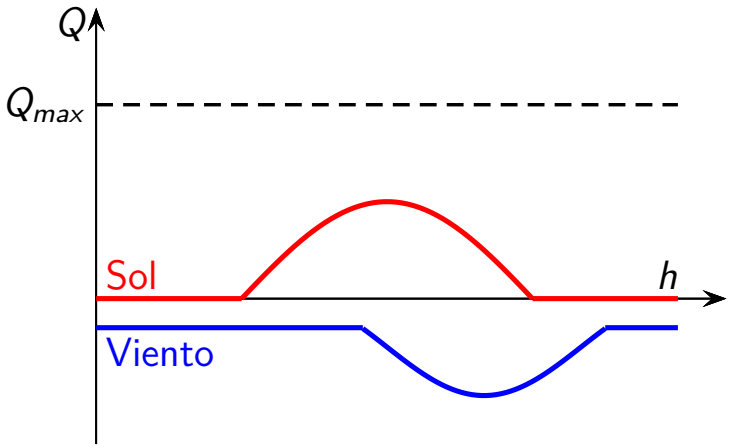
Criterio económico de diseño



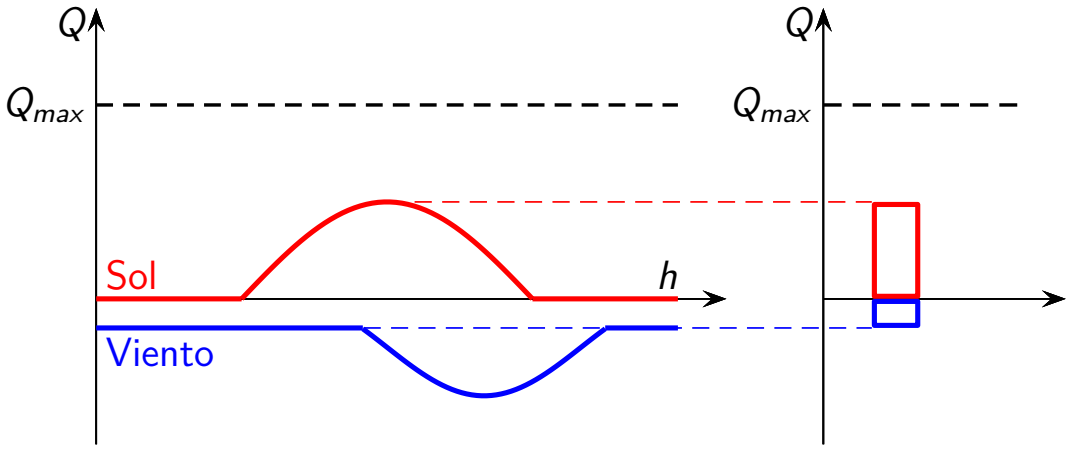
Capacidad de transporte. Diseño.



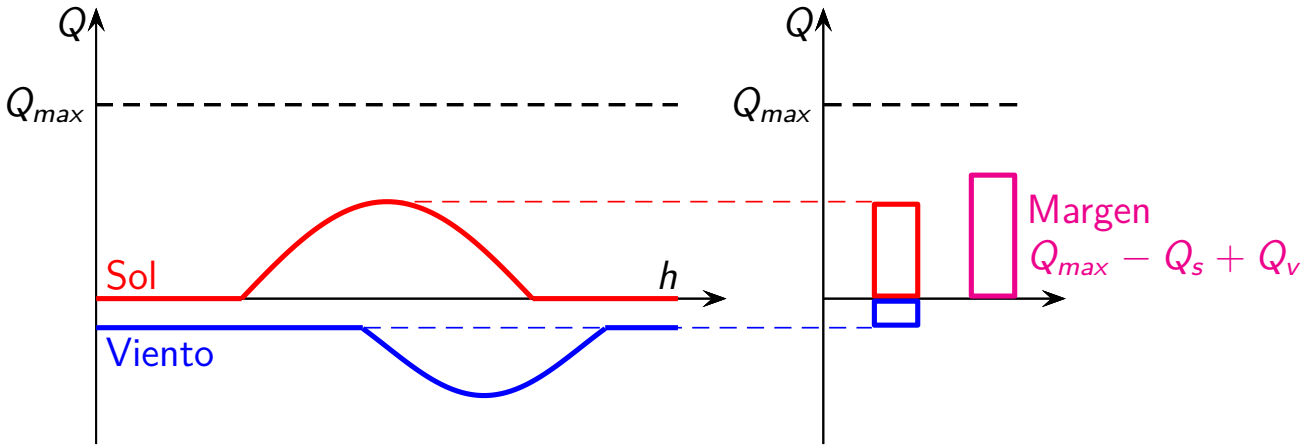
Capacidad de transporte. Diseño.



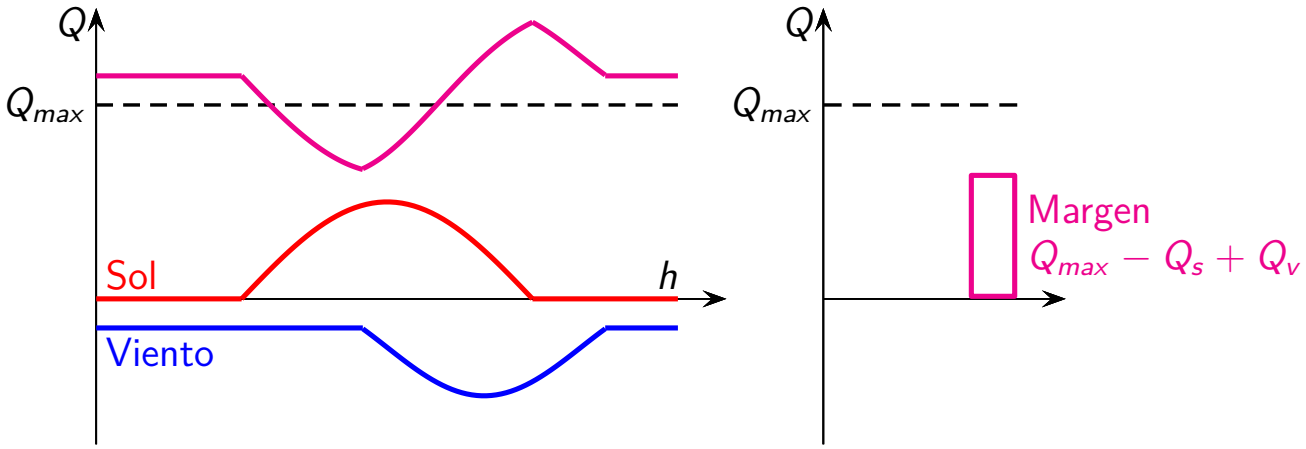
Capacidad de transporte. Diseño.



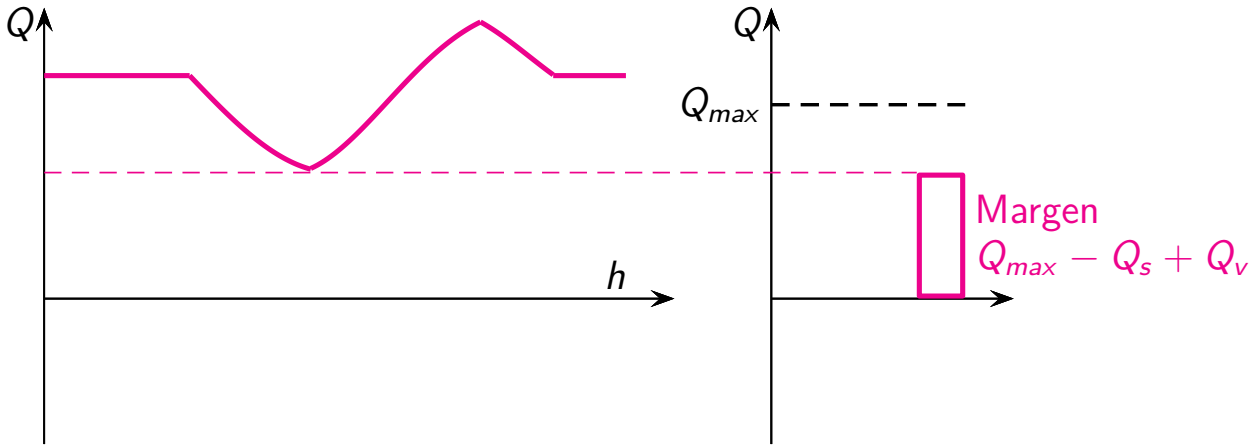
Capacidad de transporte. Diseño.



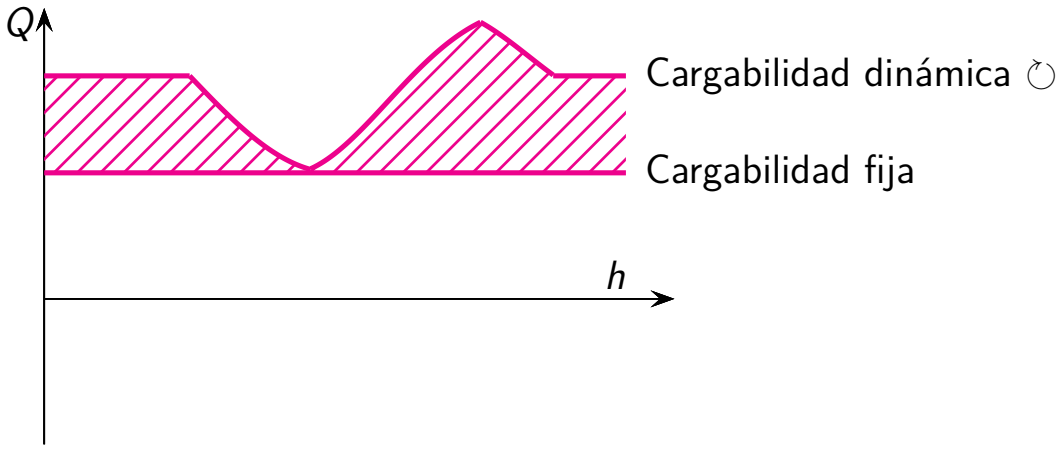
Capacidad de transporte. Diseño.



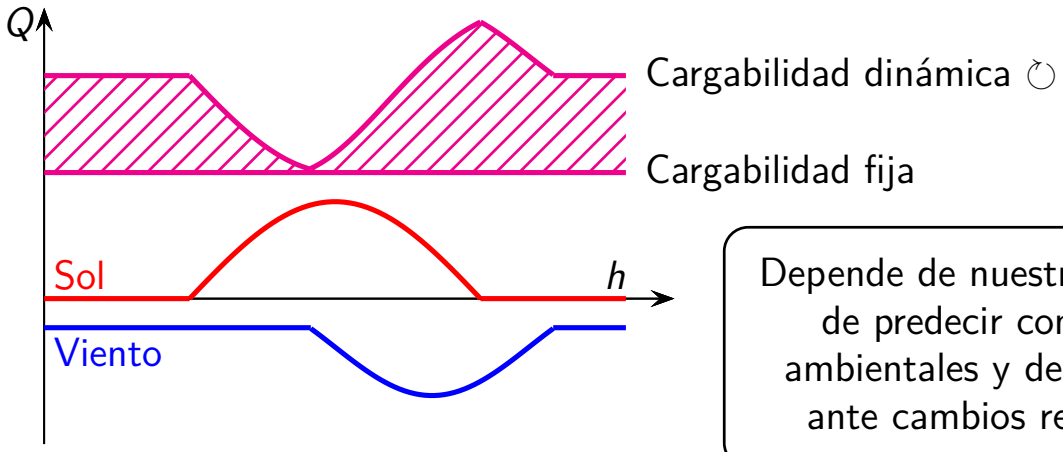
Capacidad de transporte. Diseño.



Capacidad de transporte. Diseño.
Aumento de cargabilidad con el uso de información.



Capacidad de transporte. Diseño.



Depende de nuestra capacidad de predecir condiciones ambientales y de reaccionar ante cambios repentinos.

¡Gracias!

ogduartev@unal.edu.co
@ogduartev

Ciencia de datos energéticos

Transmisión y Distribución de energía eléctrica

Oscar Duarte, Ph.D.

Universidad Nacional de Colombia
Departamento de Ingeniería Eléctrica y Electrónica

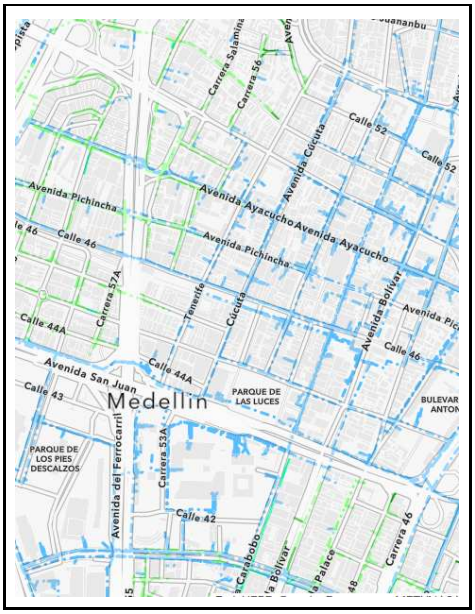
Transmisión vs Distribución
Dos problemas de transporte de energía

Transmisión

Diferencias

- Distancia
- Potencia
- Tensión
- Topología
- Control
- Medición
- Protección

Distribución

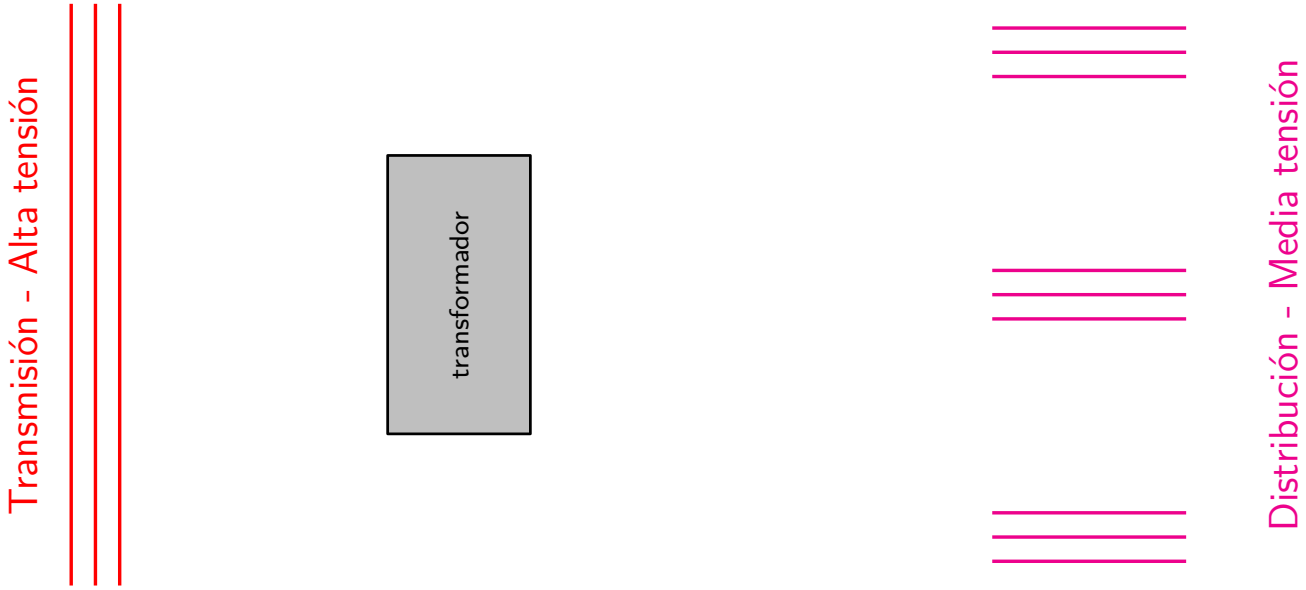


Capacidad de Transporte
Niveles de tensión, distancias y potencias

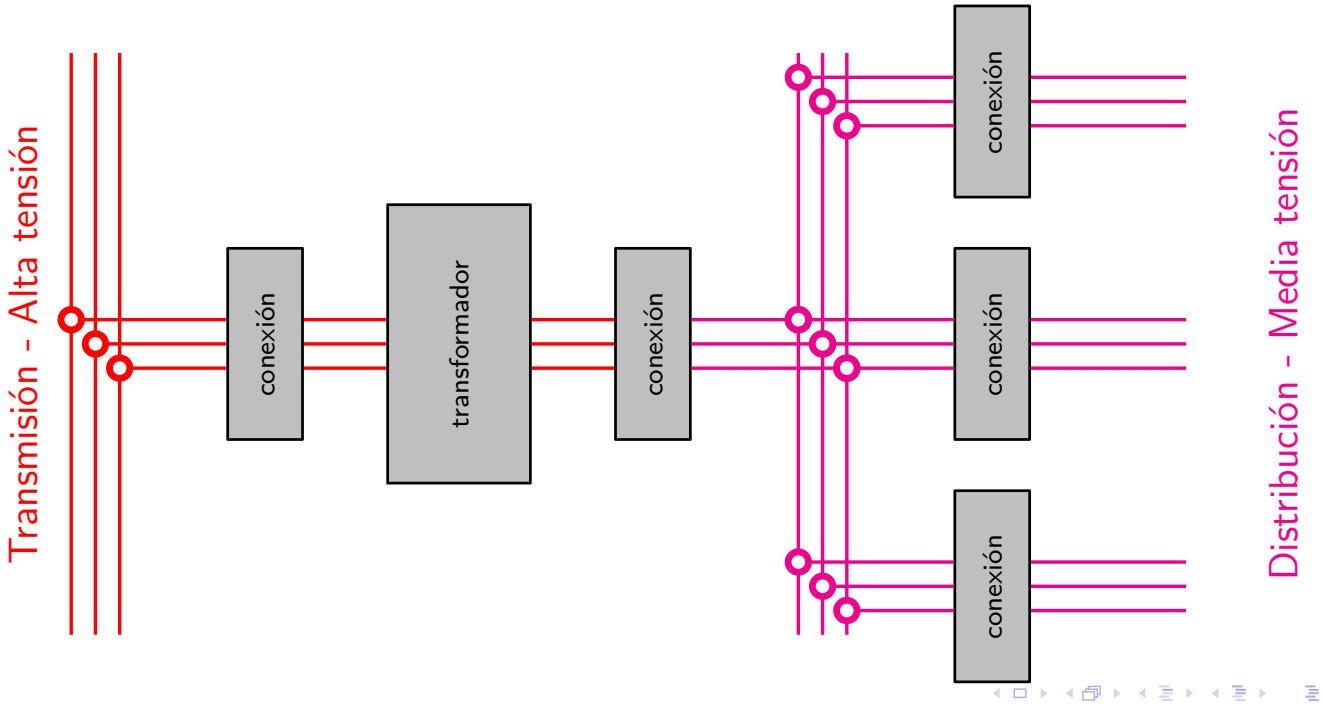
	Tensión	Distancia	Potencia
Transmisión	$> 200kV$	$> 100km$	$> 10MW$
Distribución	$< 20kV$	$\sim 10km$	$\sim 1MW$

Subestaciones
De la transmisión a la distribución

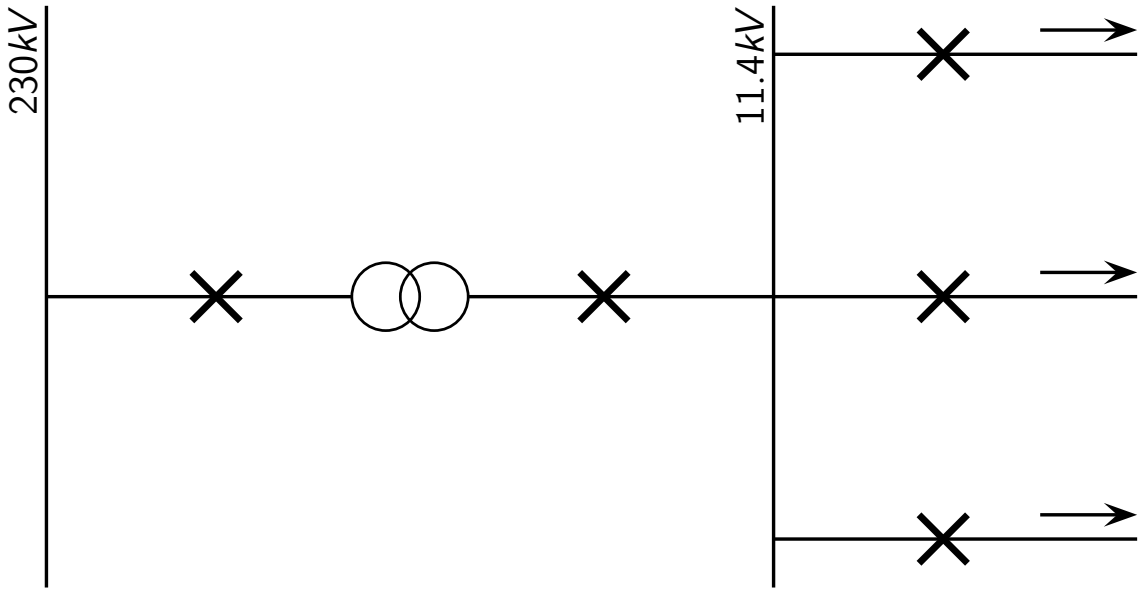
Subestaciones
De la transmisión a la distribución



Subestaciones
De la transmisión a la distribución



Subestaciones
De la transmisión a la distribución

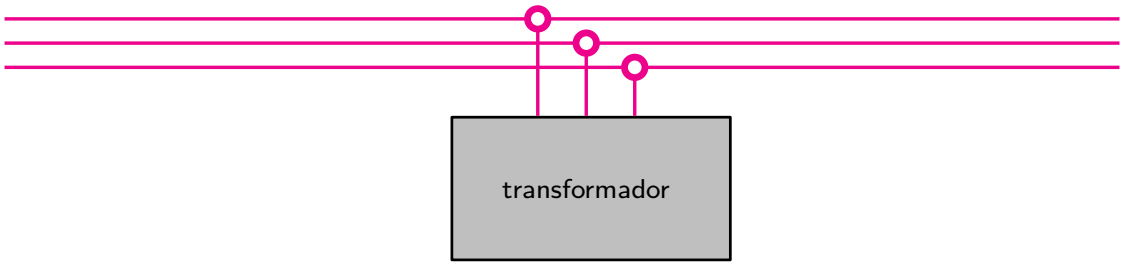


Subestaciones
De la transmisión a la distribución

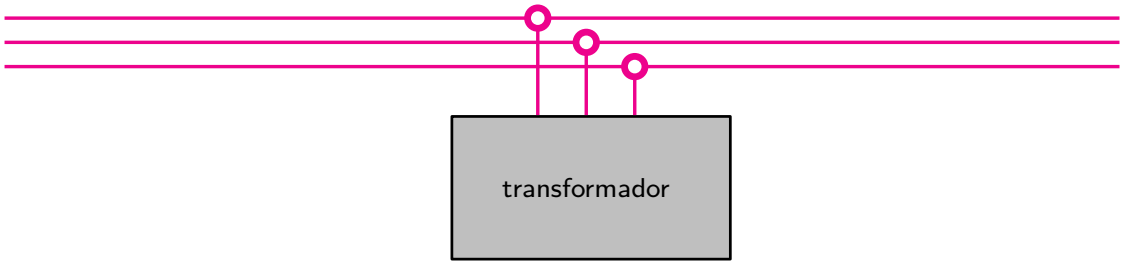
Subestaciones
De la transmisión a la distribución

Sistema de distribución
Circuitos primarios y secundarios

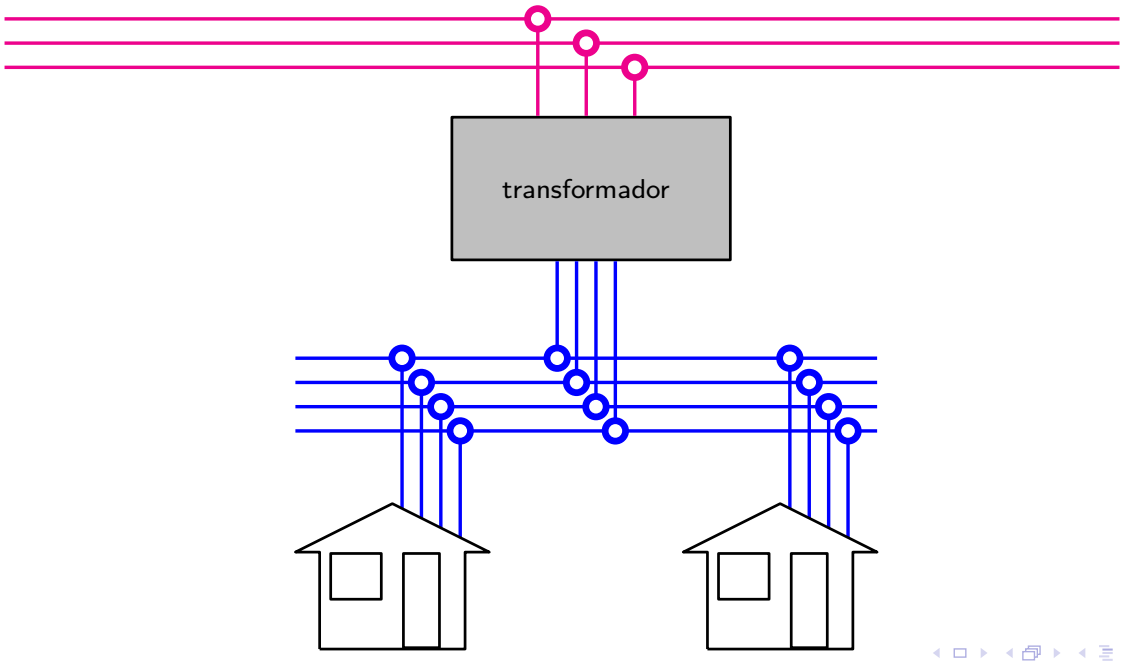
Sistema de distribución
Circuitos primarios y secundarios



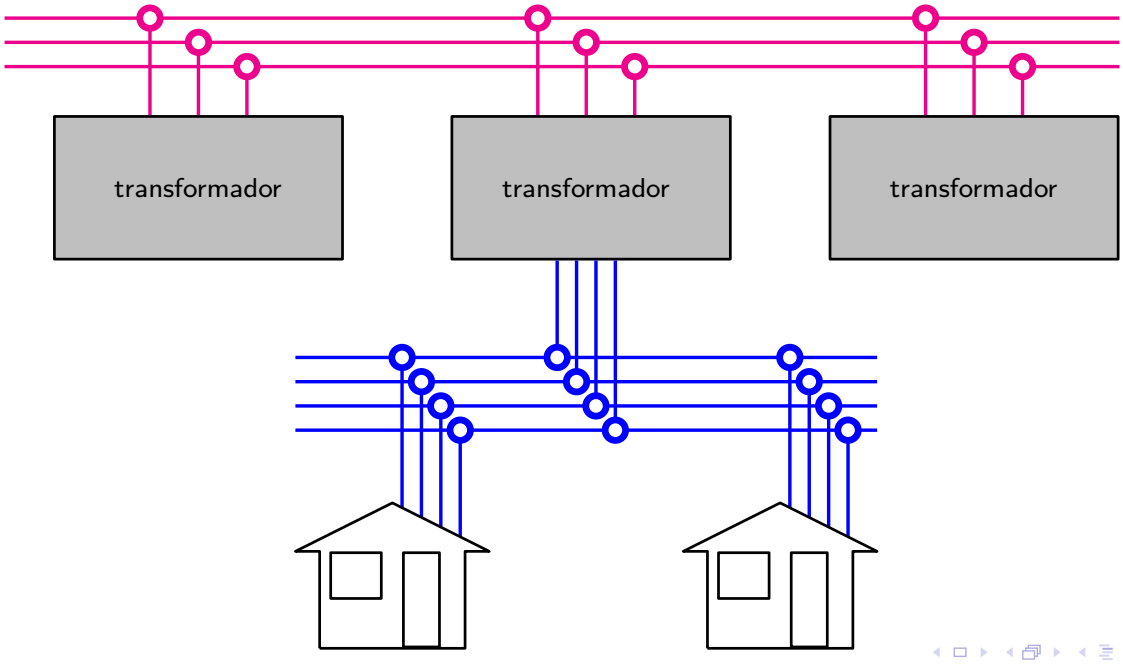
Sistema de distribución
Circuitos primarios y secundarios



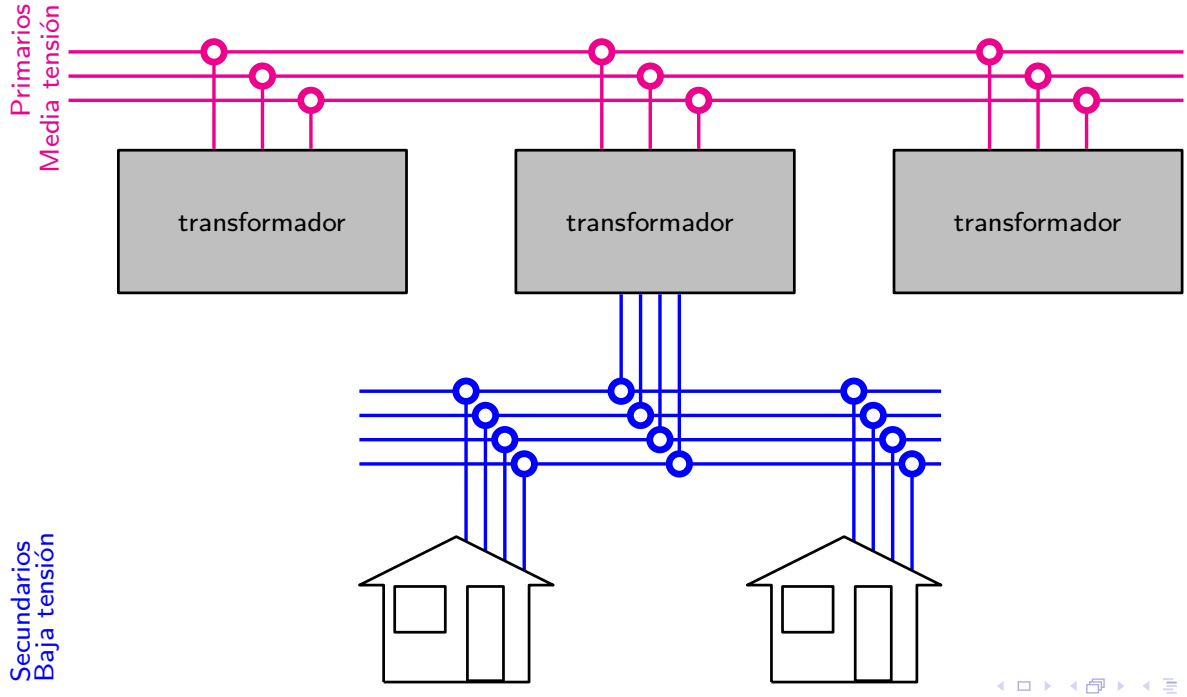
Sistema de distribución
Circuitos primarios y secundarios



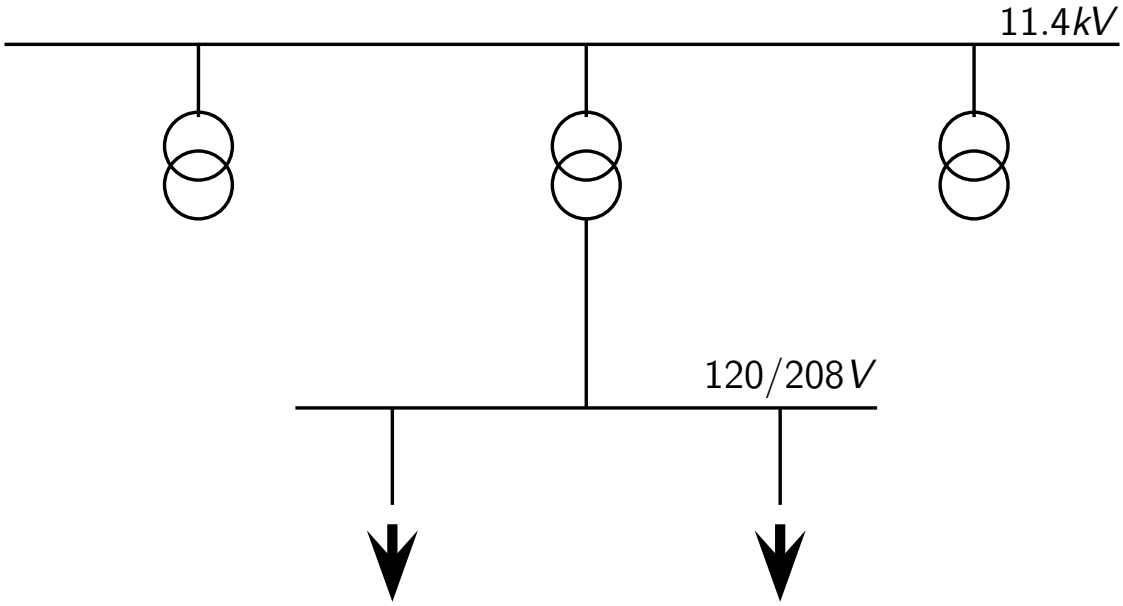
Sistema de distribución
Circuitos primarios y secundarios



Sistema de distribución
Circuitos primarios y secundarios



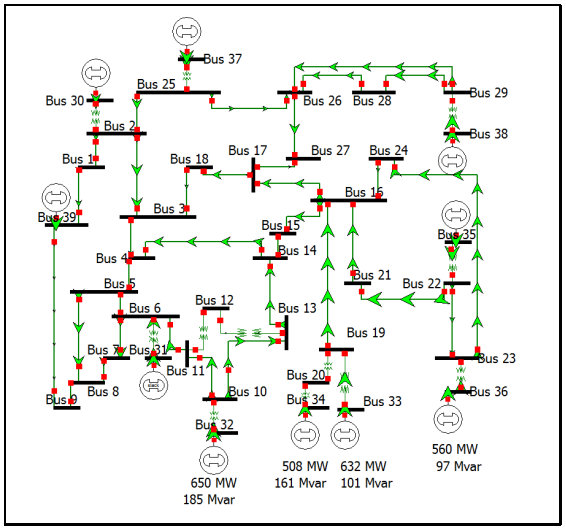
Sistema de distribución
Circuitos primarios y secundarios



Subestaciones
De la transmisión a la distribución

Transmisión vs Distribución
Topología

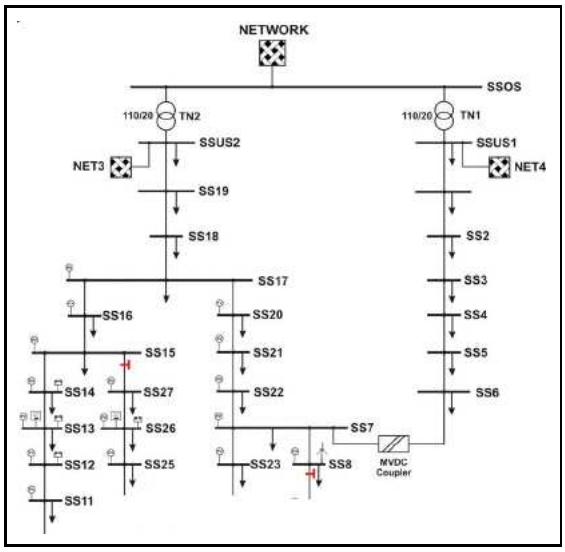
Transmisión



←Enmallada

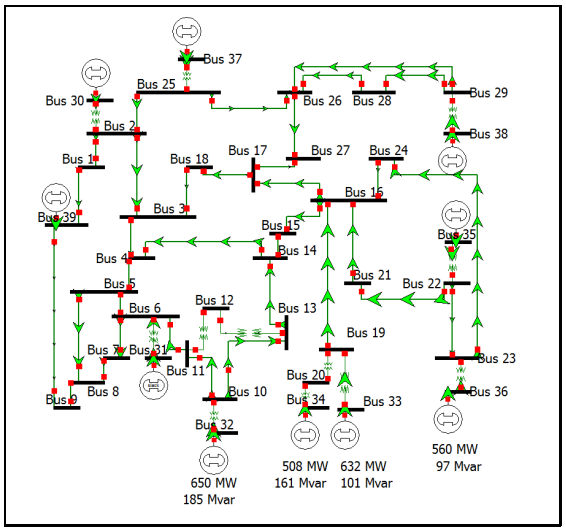
Radial→

Distribución



Transmisión vs Distribución
Control

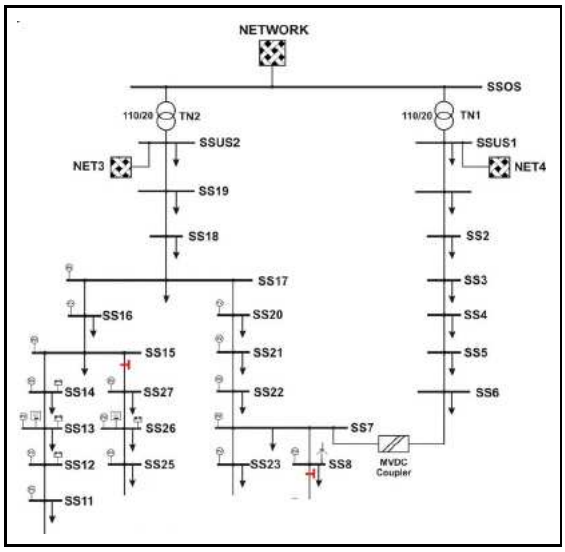
Transmisión



- Generación:
  - Potencia Activa - Frecuencia
  - Potencia Reactiva - Tensión
- Transformación:
  - Sin carga vs con carga
- Compensación reactiva
  - Sin carga vs con carga
  - Dinámica
- Conexión y desconexión
  - Protección
  - Deslastre de carga

Transmisión vs Distribución
Control

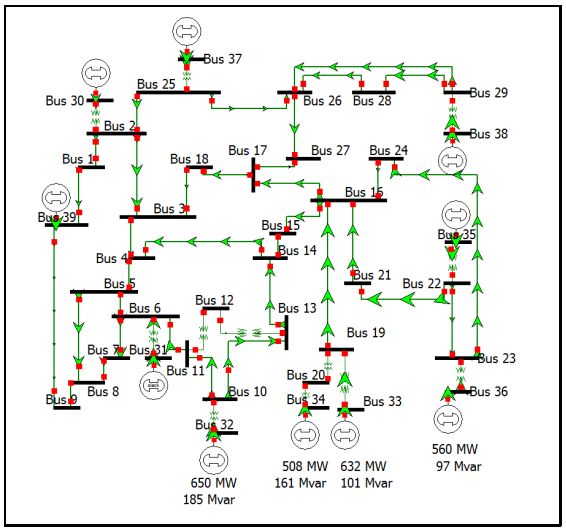
Distribución



- Transformación:
  - Solo en las subestaciones.
- Compensación reactiva
  - Líneas largas
- Conexión y desconexión
  - Protección

Transmisión vs Distribución
Mediciones

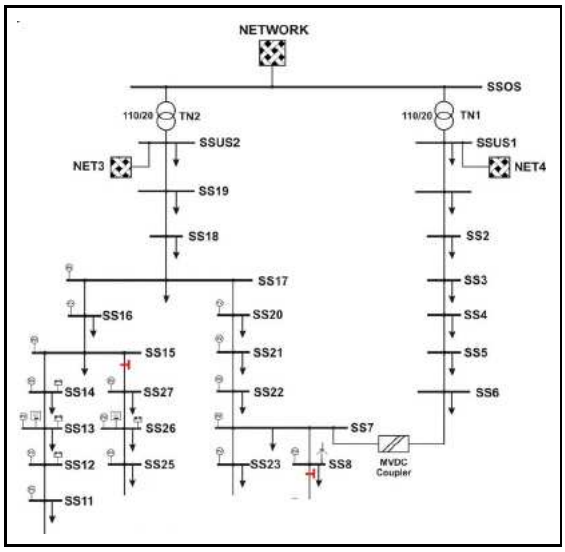
Transmisión



- En subestaciones:
  - Magnitudes de Tensiones.
  - Magnitudes de Corrientes.
  - Potencia activa, reactiva y aparente.
  - Energía.
  - Frecuencia.
  - **Unidades de medición fasorial - *PMU***
    - Fasores de Tensiones.
    - Fasores de Corrientes.
    - Frecuencia.

Transmisión vs Distribución
Mediciones

Distribución

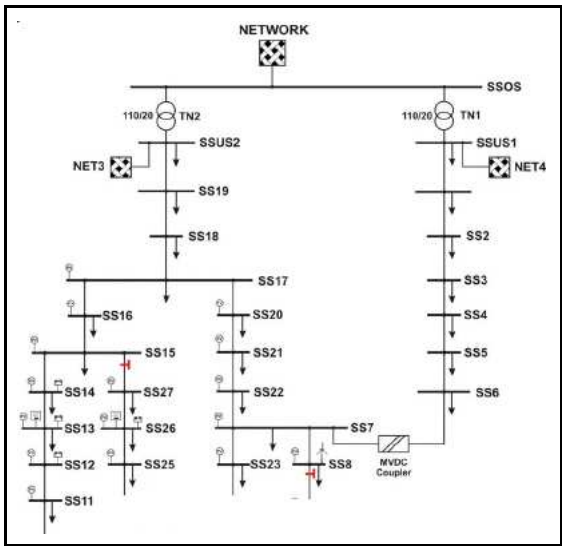


- En subestaciones:
  - Magnitudes de tensiones.
  - Magnitudes de corrientes.
  - Frecuencia.
  - Potencia activa, reactiva y aparente.
  - Energía.

Transmisión vs Distribución
Mediciones

- En los consumidores
  - Consumo mensual de energía.
  - **Instrumentos de medición Avanzada - *AMI***
    - Magnitudes de tensiones.
    - Magnitudes de corrientes.
    - Potencia activa, reactiva y aparente.
    - Energía.

Distribución



Nuevos (y viejos) retos para los sistemas de distribución
...y cómo los datos pueden ayudar a afrontarlos

- Autogeneración y flujo bidireccional de energía.
- Cargabilidad dinámica y ajuste en línea de las protecciones.
- Crecimiento no planificado de la demanda. Impacto de los vehículos eléctricos.
- Localización de fallas en sistemas subterráneos.
- Predicción de fallas y gestión de mantenimiento.

¡Gracias!

ogduartev@unal.edu.co
@ogduartev

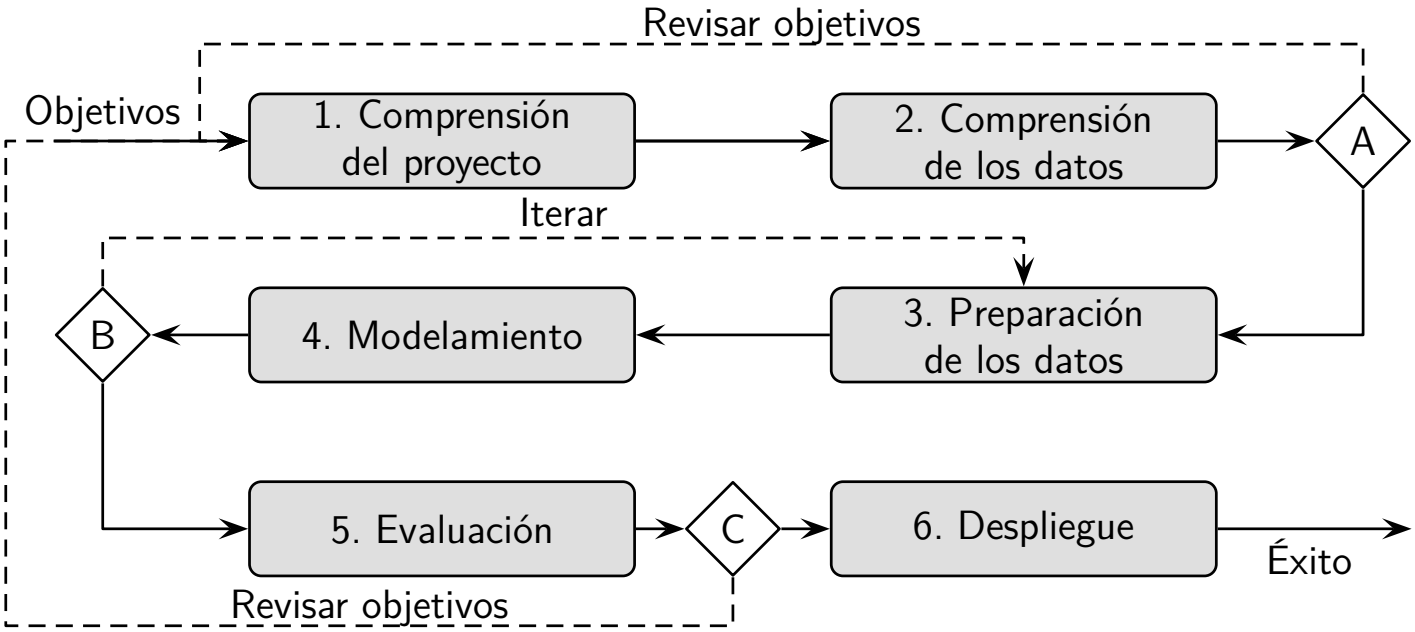
Ciencia de datos energéticos

Modelamiento de los datos - 1

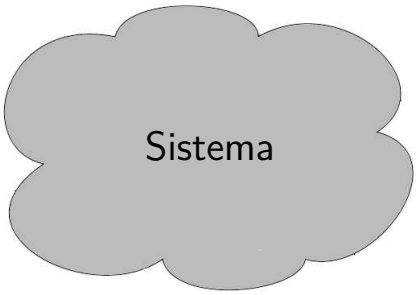
Oscar Duarte, Ph.D.

Universidad Nacional de Colombia
Departamento de Ingeniería Eléctrica y Electrónica

El proceso de la ciencia de datos
CRISP-DM: CRoss Industry Standard Process for Data Mining

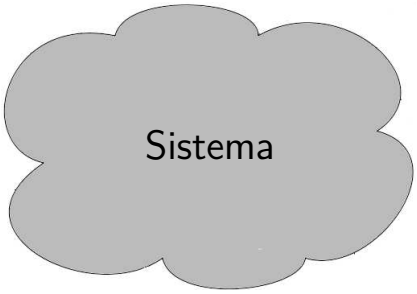


Modelamiento
Sistemas y modelos



Modelamiento
Sistemas y modelos

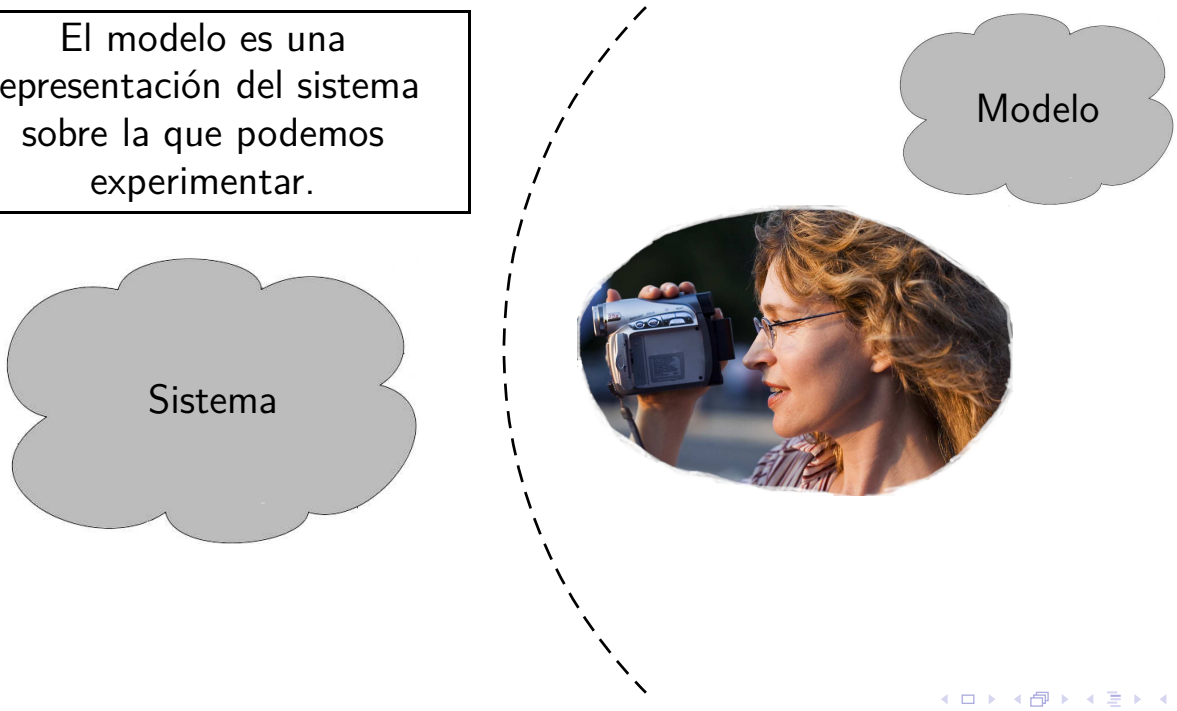
Un sistema es 'algo' bajo estudio.



Modelamiento
Sistemas y modelos

Modelamiento
Sistemas y modelos

El modelo es una representación del sistema sobre la que podemos experimentar.



Modelos

Sistema

Modelo

Modelamiento
Sobre los modelos

- Tipos de modelos: físicos, mentales, gráficos, verbales, matemáticos, de software.
- Un mismo sistema puede representarse por muchos modelos.
- ¿Cuál es el mejor modelo?
  - Utilidad.
  - Parsimonia.

Modelamiento en la ciencia de datos

Sistema

x_1	x_2	x_3	x_4

Modelamiento en la ciencia de datos

Sistema

x_1	x_2	x_3	x_4

$\Rightarrow$

Modelo

$$\mathcal{F}(\mathbf{x})$$

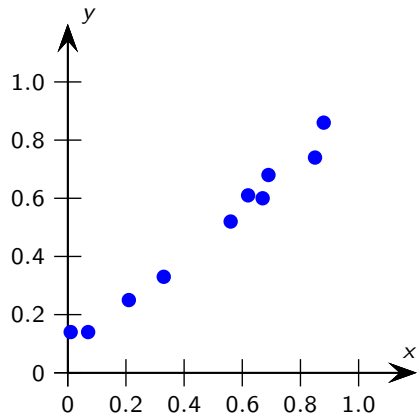
Implementación
en software

El proceso genérico

x	y
0.21	0.25
0.56	0.52
0.01	0.14
0.33	0.33
0.67	0.60
0.62	0.61
0.85	0.74
0.69	0.68
0.88	0.86
0.07	0.14

El proceso genérico

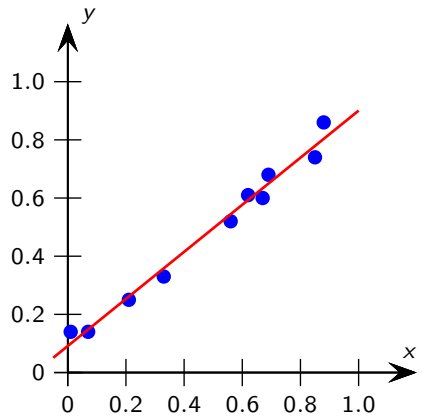
x	y
0.21	0.25
0.56	0.52
0.01	0.14
0.33	0.33
0.67	0.60
0.62	0.61
0.85	0.74
0.69	0.68
0.88	0.86
0.07	0.14



El proceso genérico

x	y
0.21	0.25
0.56	0.52
0.01	0.14
0.33	0.33
0.67	0.60
0.62	0.61
0.85	0.74
0.69	0.68
0.88	0.86
0.07	0.14

$$y = mx + b$$

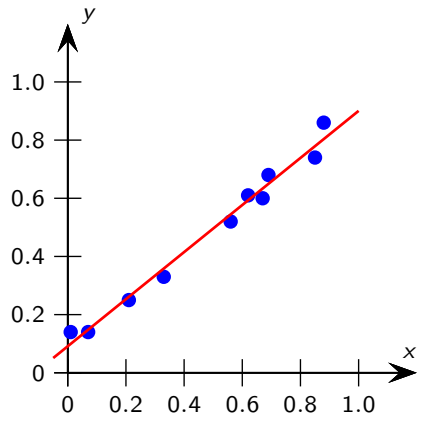


El proceso genérico

Seleccionar la
clase de modelo

x	y
0.21	0.25
0.56	0.52
0.01	0.14
0.33	0.33
0.67	0.60
0.62	0.61
0.85	0.74
0.69	0.68
0.88	0.86
0.07	0.14

$$y = mx + b$$

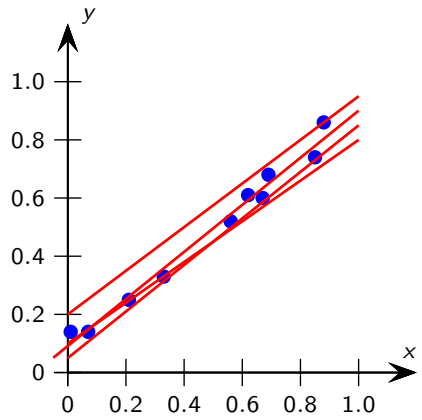


El proceso genérico

Seleccionar la
clase de modelo

x	y
0.21	0.25
0.56	0.52
0.01	0.14
0.33	0.33
0.67	0.60
0.62	0.61
0.85	0.74
0.69	0.68
0.88	0.86
0.07	0.14

¿Cuál es la mejor?



El proceso genérico

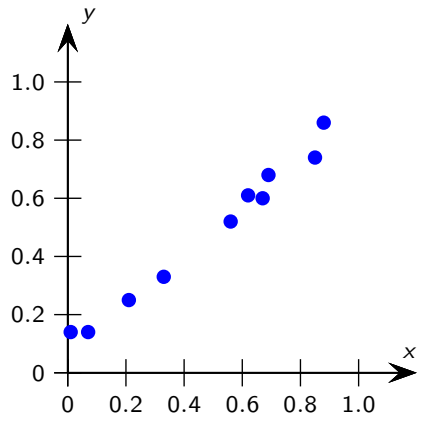
Seleccionar la
clase de modelo

↓

Seleccionar la
función de error

x	y
0.21	0.25
0.56	0.52
0.01	0.14
0.33	0.33
0.67	0.60
0.62	0.61
0.85	0.74
0.69	0.68
0.88	0.86
0.07	0.14

Mínimo error
cuadrático medio



El proceso genérico

Seleccionar la **clase de modelo**

↓

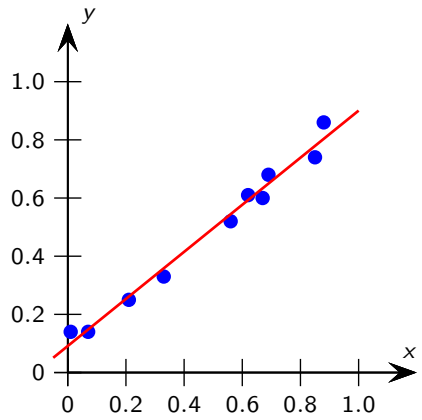
Seleccionar la **función de error**

↓

Aplicar el **algoritmo**

x	y
0.21	0.25
0.56	0.52
0.01	0.14
0.33	0.33
0.67	0.60
0.62	0.61
0.85	0.74
0.69	0.68
0.88	0.86
0.07	0.14

$$[m\ b]^T = \mathbf{y}/\mathbf{X}$$



El proceso genérico

Seleccionar la **clase de modelo**

↓

Seleccionar la **función de error**

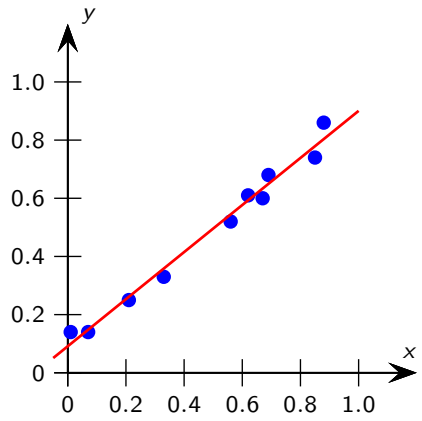
↓

Aplicar el **algoritmo**

↓

Validar el resultado

x	y
0.21	0.25
0.56	0.52
0.01	0.14
0.33	0.33
0.67	0.60
0.62	0.61
0.85	0.74
0.69	0.68
0.88	0.86
0.07	0.14



El proceso genérico

Seleccionar la **clase de modelo**

↓

Seleccionar la **función de error**

↓

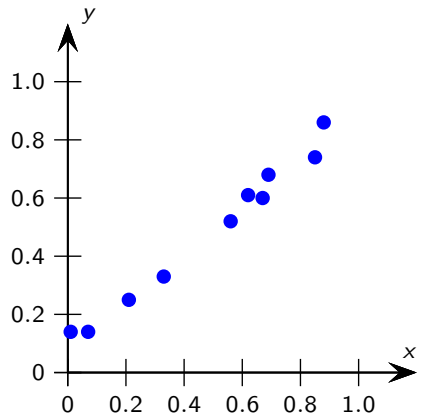
Aplicar el **algoritmo**

↓

Validar el resultado

x	y
0.21	0.25
0.56	0.52
0.01	0.14
0.33	0.33
0.67	0.60
0.62	0.61
0.85	0.74
0.69	0.68
0.88	0.86
0.07	0.14

¿Qué tal un polinomio?



El proceso genérico

Seleccionar la **clase de modelo**

↓

Seleccionar la **función de error**

↓

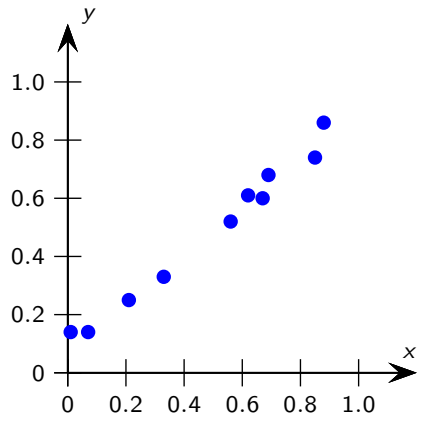
Aplicar el **algoritmo**

↓

Validar el resultado

Orden	Error (%)
0.	37.25
1.	3.34
2.	2.29
3.	1.91
4.	1.50
5.	1.98
6.	1.10
7.	0.82
8.	0.82
9.	0.00

¿Qué tal un polinomio?



El proceso genérico

Seleccionar la **clase de modelo**

↓

Seleccionar la **función de error**

↓

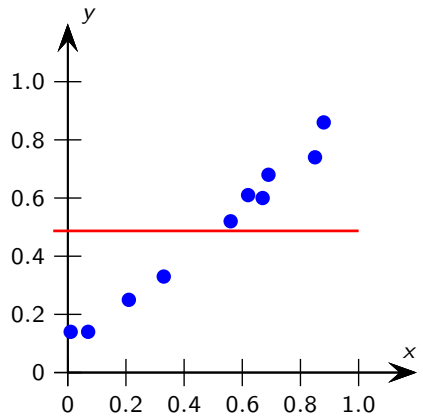
Aplicar el **algoritmo**

↓

Validar el resultado

Orden	Error (%)
0.	37.25
1.	3.34
2.	2.29
3.	1.91
4.	1.50
5.	1.98
6.	1.10
7.	0.82
8.	0.82
9.	0.00

Orden 0



El proceso genérico

Seleccionar la **clase de modelo**

↓

Seleccionar la **función de error**

↓

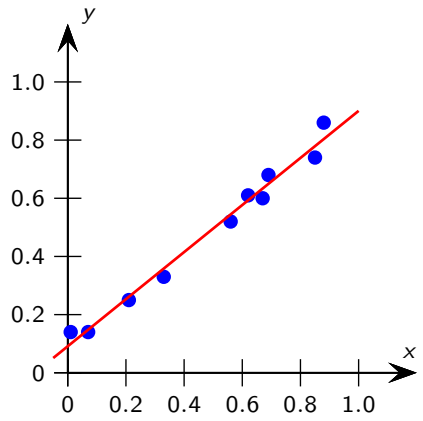
Aplicar el **algoritmo**

↓

Validar el resultado

Orden	Error (%)
0.	37.25
1.	3.34
2.	2.29
3.	1.91
4.	1.50
5.	1.98
6.	1.10
7.	0.82
8.	0.82
9.	0.00

Orden 1



El proceso genérico

Seleccionar la **clase de modelo**

↓

Seleccionar la **función de error**

↓

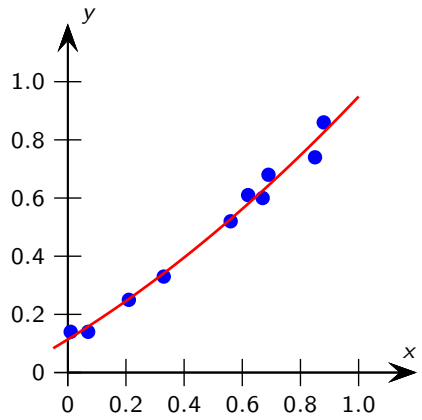
Aplicar el **algoritmo**

↓

Validar el resultado

Orden	Error (%)
0.	37.25
1.	3.34
2.	2.29
3.	1.91
4.	1.50
5.	1.98
6.	1.10
7.	0.82
8.	0.82
9.	0.00

Orden 2



El proceso genérico

Seleccionar la **clase de modelo**

↓

Seleccionar la **función de error**

↓

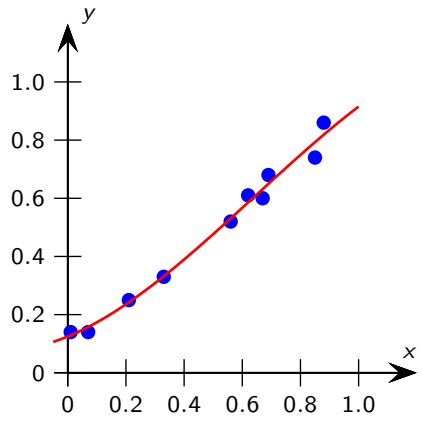
Aplicar el **algoritmo**

↓

Validar el resultado

Orden	Error (%)
0.	37.25
1.	3.34
2.	2.29
3.	1.91
4.	1.50
5.	1.98
6.	1.10
7.	0.82
8.	0.82
9.	0.00

Orden 3



El proceso genérico

Seleccionar la **clase de modelo**

↓

Seleccionar la **función de error**

↓

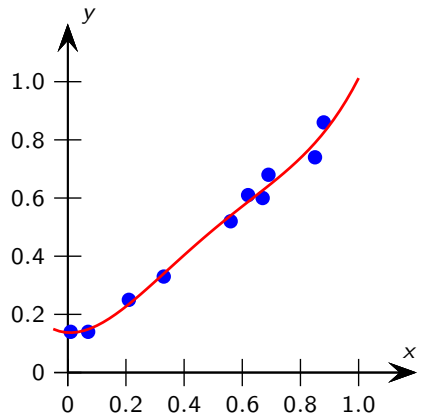
Aplicar el **algoritmo**

↓

Validar el resultado

Orden	Error (%)
0.	37.25
1.	3.34
2.	2.29
3.	1.91
4.	1.50
5.	1.98
6.	1.10
7.	0.82
8.	0.82
9.	0.00

Orden 4



El proceso genérico

Seleccionar la **clase de modelo**

↓

Seleccionar la **función de error**

↓

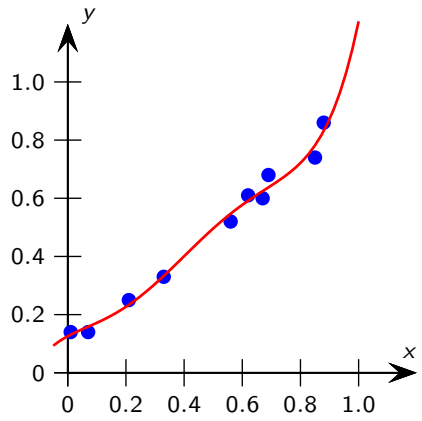
Aplicar el **algoritmo**

↓

Validar el resultado

Orden	Error (%)
0.	37.25
1.	3.34
2.	2.29
3.	1.91
4.	1.50
5.	1.98
6.	1.10
7.	0.82
8.	0.82
9.	0.00

Orden 5



El proceso genérico

Seleccionar la **clase de modelo**

↓

Seleccionar la **función de error**

↓

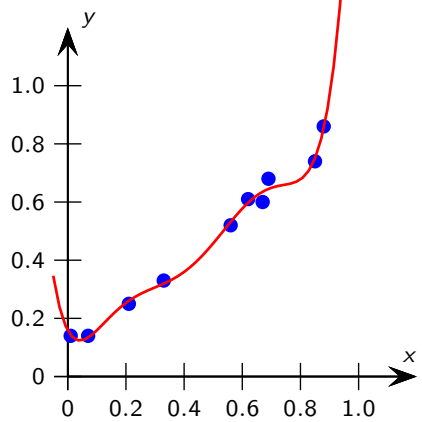
Aplicar el **algoritmo**

↓

Validar el resultado

Orden	Error (%)
0.	37.25
1.	3.34
2.	2.29
3.	1.91
4.	1.50
5.	1.98
6.	1.10
7.	0.82
8.	0.82
9.	0.00

Orden 6



El proceso genérico

Seleccionar la **clase de modelo**

↓

Seleccionar la **función de error**

↓

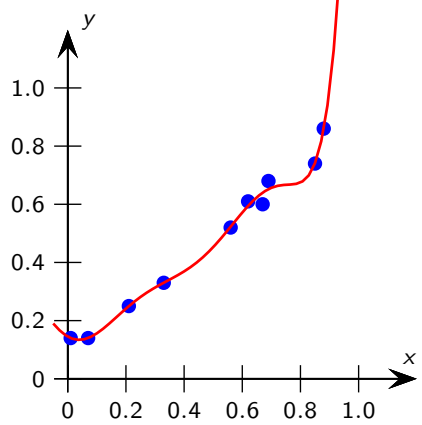
Aplicar el **algoritmo**

↓

Validar el resultado

Orden	Error (%)
0.	37.25
1.	3.34
2.	2.29
3.	1.91
4.	1.50
5.	1.98
6.	1.10
7.	0.82
8.	0.82
9.	0.00

Orden 7



El proceso genérico

Seleccionar la **clase de modelo**

↓

Seleccionar la **función de error**

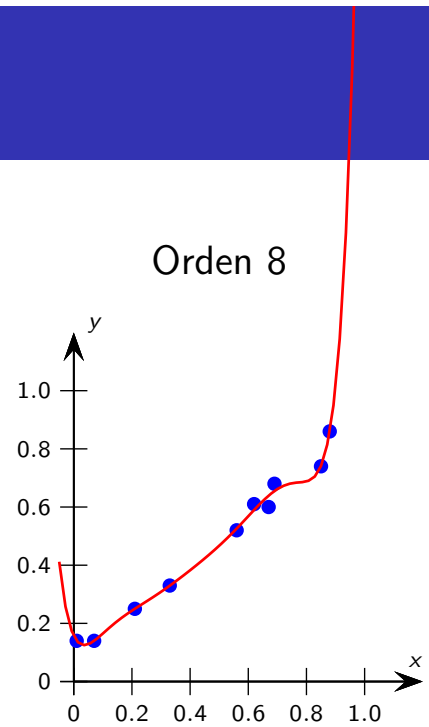
↓

Aplicar el **algoritmo**

↓

Validar el resultado

Orden	Error (%)
0.	37.25
1.	3.34
2.	2.29
3.	1.91
4.	1.50
5.	1.98
6.	1.10
7.	0.82
8.	0.82
9.	0.00



El proceso genérico

Seleccionar la **clase de modelo**

↓

Seleccionar la **función de error**

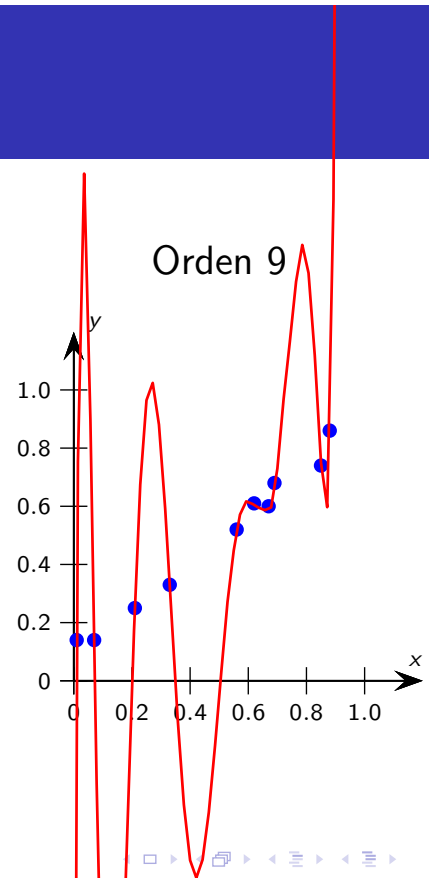
↓

Aplicar el **algoritmo**

↓

Validar el resultado

Orden	Error (%)
0.	37.25
1.	3.34
2.	2.29
3.	1.91
4.	1.50
5.	1.98
6.	1.10
7.	0.82
8.	0.82
9.	0.00



El proceso genérico

Seleccionar la **clase de modelo**

↓

Seleccionar la **función de error**

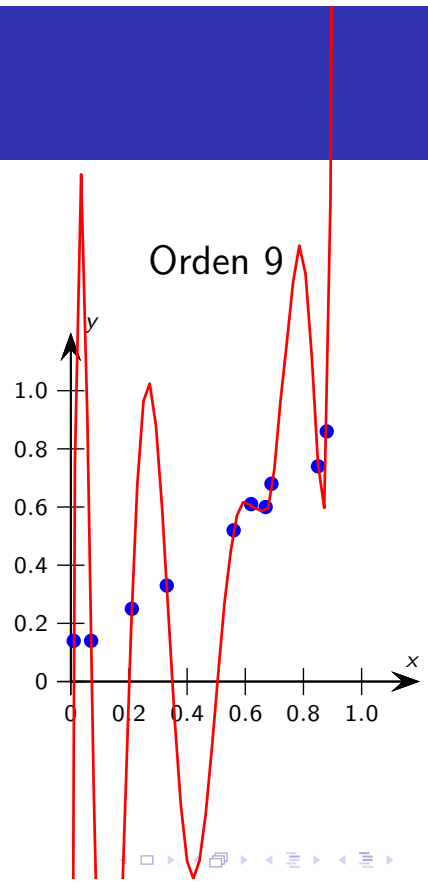
↓

Aplicar el **algoritmo**

↓

Validar el resultado

- Generalización.
- Parsimonia.
- Interpretabilidad.
- Precisión.



¡Gracias!

ogduartev@unal.edu.co
@ogduartev

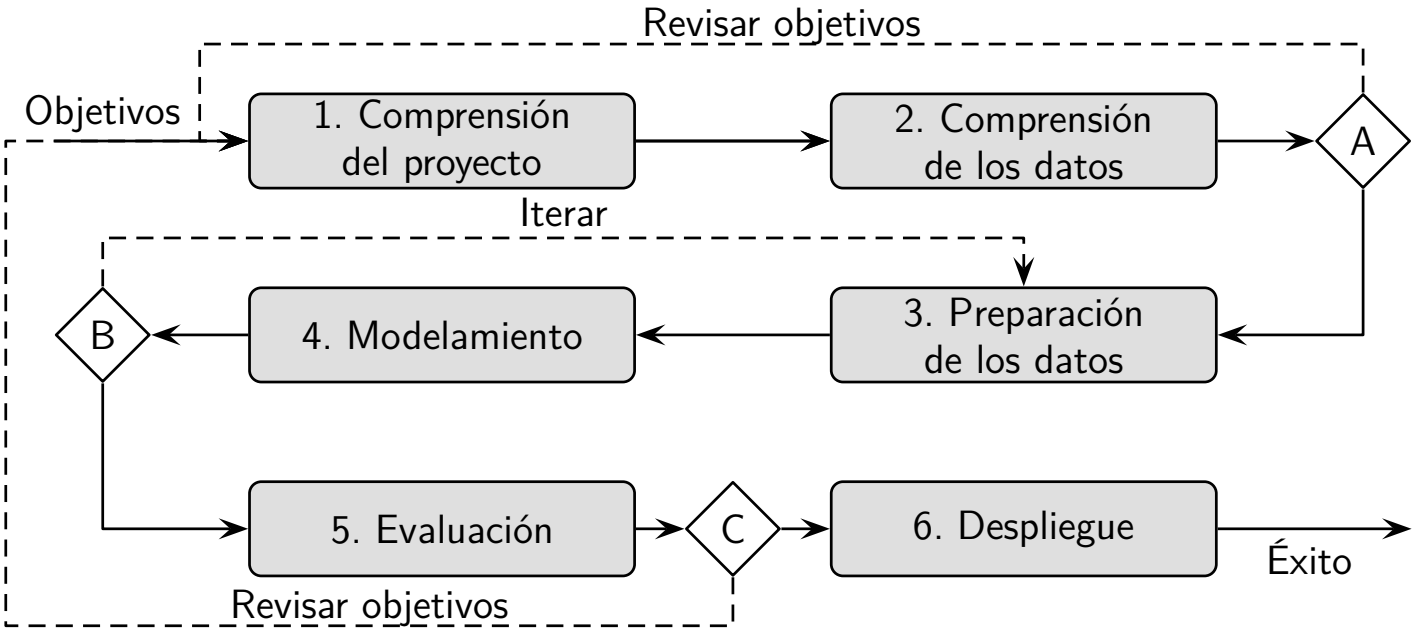
Ciencia de datos energéticos

Modelamiento de los datos - 2

Oscar Duarte, Ph.D.

Universidad Nacional de Colombia
Departamento de Ingeniería Eléctrica y Electrónica

El proceso de la ciencia de datos
CRISP-DM: CRoss Industry Standard Process for Data Mining



Modelamiento en la ciencia de datos

Sistema

x_1	x_2	x_3	x_4

$\Rightarrow$

Modelo

$$\mathcal{F}(\mathbf{x})$$

Implementación
en software

Funciones de error
El objetivo es minimizar el *error*

$f : \mathcal{M} \to \mathbb{R}$ $\qquad$ f es la *Función Objetivo*

$\mathcal{M}$ = Conjunto de todos los modelos de la clase seleccionada

Funciones de error
El objetivo es minimizar el *error*

$$f : \mathcal{M} \to \mathbb{R} \qquad f \text{ es la } \textit{Función Objetivo}$$

$$\begin{aligned} \mathcal{M} &= \text{Conjunto de todos los modelos de la clase seleccionada} \\ &= \text{Conjunto de todas las rectas } y = mx + b \end{aligned}$$

Funciones de error
El objetivo es minimizar el *error*

$$f : \mathcal{M} \to \mathbb{R} \qquad f \text{ es la } \textit{Función Objetivo}$$

$$\begin{aligned} \mathcal{M} &= \text{Conjunto de todos los modelos de la clase seleccionada} \\ &= \text{Conjunto de todas las rectas } y = mx + b \end{aligned}$$

$$= \text{Conjunto de todas las combinaciones posibles de } m, b$$

Funciones de error
El objetivo es minimizar el *error*

$$f : \mathcal{M} \to \mathbb{R} \qquad f \text{ es la } \textit{Función Objetivo}$$

$$\begin{aligned}
\mathcal{M} &= \text{Conjunto de todos los modelos de la clase seleccionada} \\
&= \text{Conjunto de todas las rectas } y = mx + b \\
&= \text{Conjunto de todas las combinaciones posibles de } m, b \\
&= \{(m, b) \mid m \in \mathbb{R},\ b \in \mathbb{R}\}
\end{aligned}$$

Funciones de error
El objetivo es minimizar el *error*

$$f : \mathcal{M} \to \mathbb{R} \qquad f \text{ es la } \textit{Función Objetivo}$$

$$\begin{aligned}
\mathcal{M} &= \text{Conjunto de todos los modelos de la clase seleccionada} \\
&= \text{Conjunto de todas las rectas } y = mx + b \\
&= \text{Conjunto de todas las combinaciones posibles de } m, b \\
&= \{(m, b) \mid m \in \mathbb{R},\ b \in \mathbb{R}\}
\end{aligned}$$

$$= \mathbb{R}^2$$

Funciones de error
El objetivo es minimizar el *error*

$$f : \mathcal{M} \to \mathbb{R} \qquad f \text{ es la } \textit{Función Objetivo}$$

$$\begin{aligned}
\mathcal{M} &= \text{Conjunto de todos los modelos de la clase seleccionada} \\
&= \text{Conjunto de todas las rectas } y = mx + b \\
&= \text{Conjunto de todas las combinaciones posibles de } m, b \\
&= \{(m, b) \mid m \in \mathbb{R},\ b \in \mathbb{R}\} \\
&= \mathbb{R}^2
\end{aligned}$$

$$f : \mathbb{R}^2 \to \mathbb{R}$$

Funciones de error
El objetivo es minimizar el *error*

$$f : \mathbb{R}^2 \to \mathbb{R} \qquad f \text{ es la } \textit{Función Objetivo}$$

Funciones de error
El objetivo es minimizar el *error*

$f : \mathbb{R}^2 \to \mathbb{R}$ $\qquad$ f es la *Función Objetivo*

x	**y**	$\tilde{\mathbf{y}}$	**e**
x_1	y_1	$\tilde{y}_1$	e_1
x_2	y_2	$\tilde{y}_2$	e_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	y_n	$\tilde{y}_n$	e_n

$$\tilde{y}_i = mx_i + b$$
$$e = \tilde{y}_i - y_i$$

Funciones de error
El objetivo es minimizar el *error*

$f : \mathbb{R}^2 \to \mathbb{R}$ $\qquad$ f es la *Función Objetivo*

x	**y**	$\tilde{\mathbf{y}}$	**e**
x_1	y_1	$\tilde{y}_1$	e_1
x_2	y_2	$\tilde{y}_2$	e_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	y_n	$\tilde{y}_n$	e_n

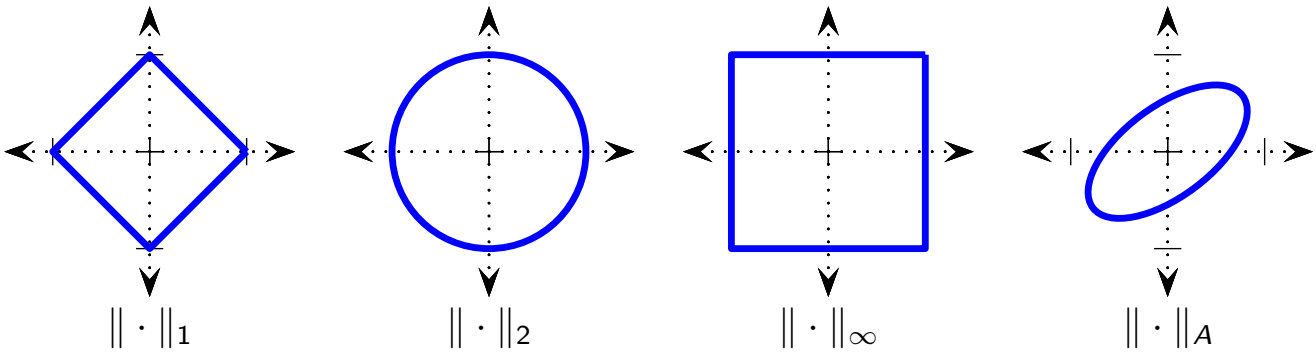
$$\tilde{y}_i = mx_i + b$$
$$e = \tilde{y}_i - y_i$$

$$\varepsilon = \frac{1}{n}\sqrt{\sum_{i=1}^{n} {e_i}^2}$$

$$\varepsilon = \frac{1}{n}\sqrt{\sum_{i=1}^{n} (mx_i + b - y_i)^2}$$

El error es una distancia
Circunferencia unitaria para distintas Normas

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{q} x_i^p\right)^{1/p} \qquad q \text{ dimensiones}$$



$$A = \begin{bmatrix} 2.08 & -1.44 \\ -1.44 & 2.92 \end{bmatrix}$$

Errores de clasificación

Matriz de costos

Clase Verdadera	Clase predicha			
	C_1	C_2	$\cdots$	C_m
C_1	0	$c_{1,2}$	$\cdots$	$c_{1,m}$
C_2	$c_{2,1}$	0	$\cdots$	$c_{2,m}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
C_m	$c_{m,1}$	$c_{m,2}$	$\cdots$	0

Errores de clasificación

Matriz de costos

Clase Verdadera	Clase predicha			
	C_1	C_2	$\cdots$	C_m
C_1	0	$c_{1,2}$	$\cdots$	$c_{1,m}$
C_2	$c_{2,1}$	0	$\cdots$	$c_{2,m}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
C_m	$c_{m,1}$	$c_{m,2}$	$\cdots$	0

$$\text{loss}\,(C_i|E) = \sum_{i=1}^{m} P(C_j|E)c_{j,i}$$

Errores de clasificación
Dos clases

Matriz de confusión

Clase Verdadera	Clase predicha	
	P	N
P		
N		

Errores de clasificación
Dos clases

Matriz de confusión

Clase Verdadera	Clase predicha	
	P	N
P	VP	FN
N	FP	VN

Errores de clasificación
Dos clases

Matriz de confusión

Clase Verdadera	Clase predicha	
	P	N
P	VP	FN
N	FP	VN

$$\text{Recall} = \frac{VP}{VP + FN}$$

$$\text{Especificidad} = \frac{VN}{VN + FP}$$

$$\text{Precisión} = \frac{VP}{VP + FP}$$

$$\text{Exactitud} = \frac{VP + VN}{VP + FN + FP + VN}$$

Errores de clasificación
Dos clases

Matriz de confusión

Clase Verdadera	Clase predicha	
	P	N
P	VP	FN
N	FP	VN

$$F = 2\frac{\text{Precisión} \times \text{Recall}}{\text{Precisión} + \text{Recall}}$$

$$\text{Recall} = \frac{VP}{VP + FN}$$

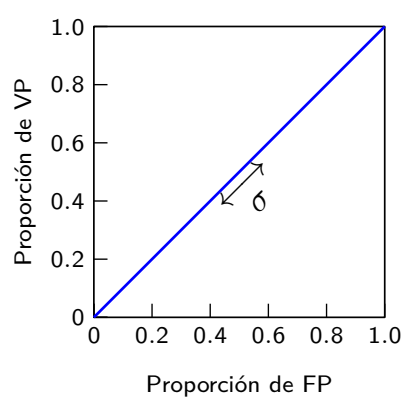
$$\text{Especificidad} = \frac{VN}{VN + FP}$$

$$\text{Precisión} = \frac{VP}{VP + FP}$$

$$\text{Exactitud} = \frac{VP + VN}{VP + FN + FP + VN}$$

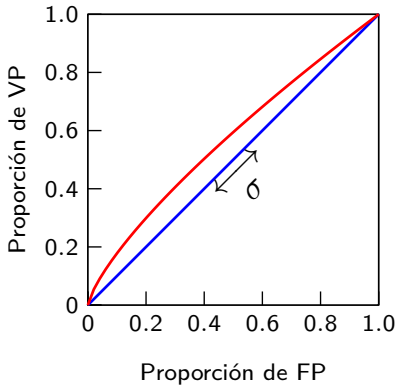
Errores de clasificación
Curva ROC (*Receiver Operating Characteristic*)

Clasificador
Parámetro σ



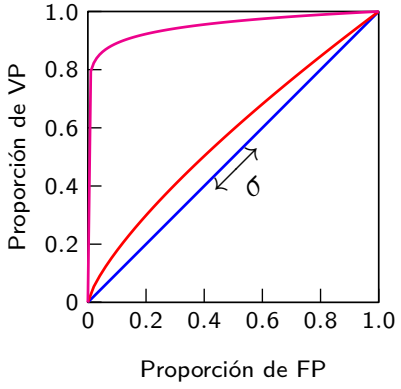
Errores de clasificación
Curva ROC (*Receiver Operating Characteristic*)

Clasificador
Parámetro σ



Errores de clasificación
Curva ROC (*Receiver Operating Characteristic*)

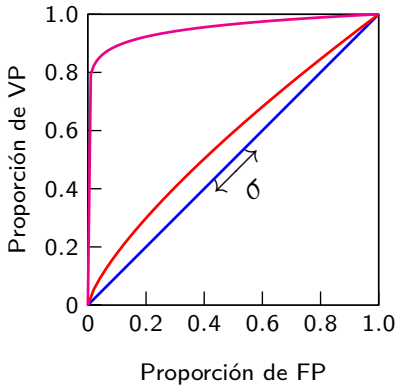
Clasificador
Parámetro σ



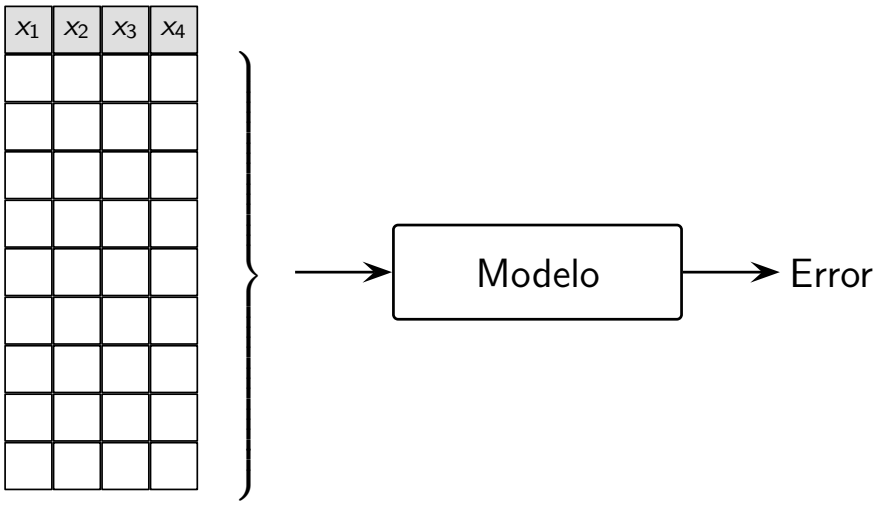
Errores de clasificación
Curva ROC (*Receiver Operating Characteristic*)

AUC: Área Bajo la Curva ROC

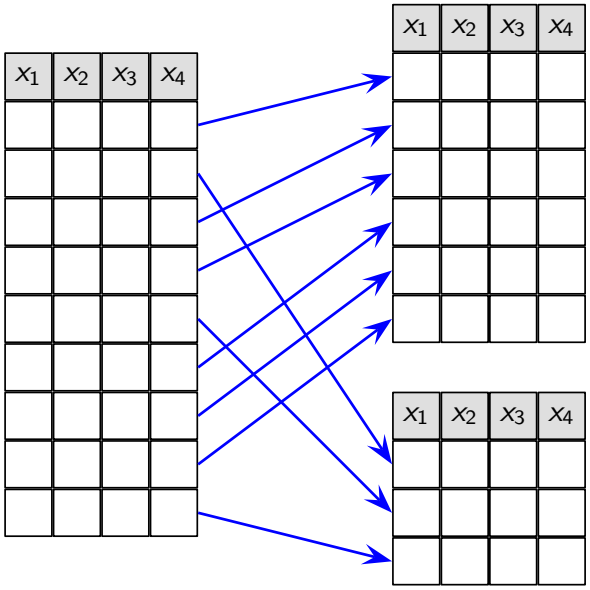
Clasificador
Parámetro σ



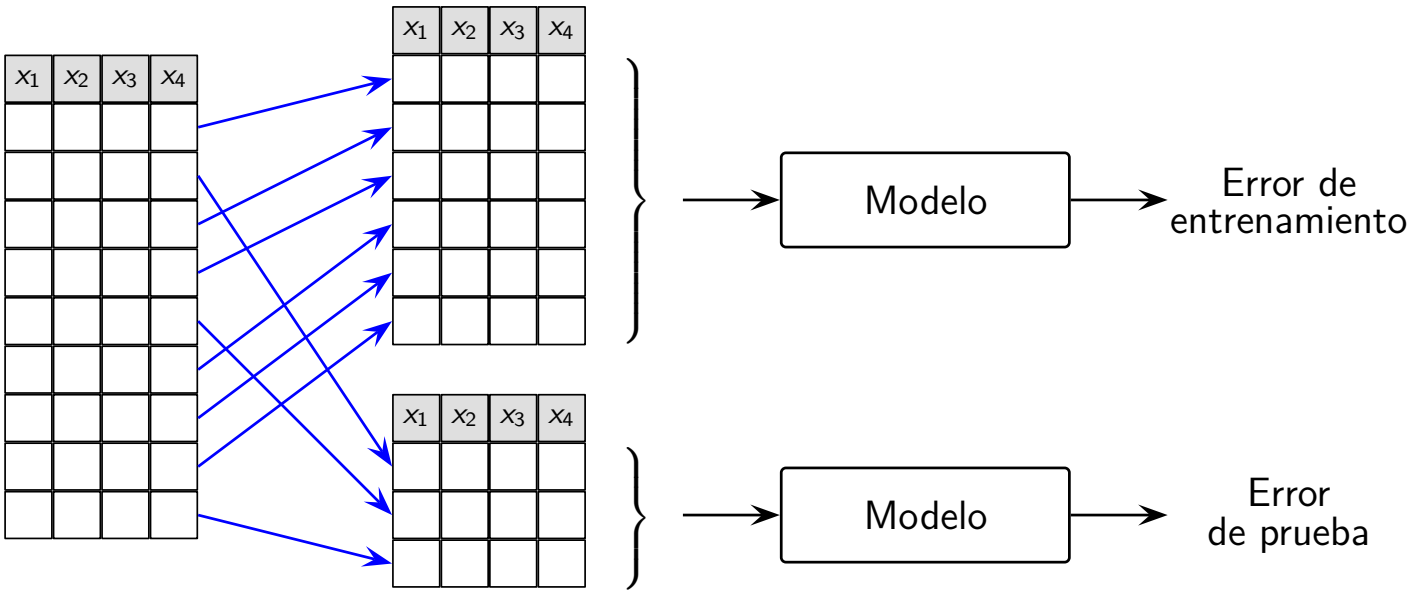
Estrategias de validación



Estrategias de validación



Estrategias de validación
Subentrenamiento vs Sobreentrenamiento



¡Gracias!

ogduartev@unal.edu.co
@ogduartev

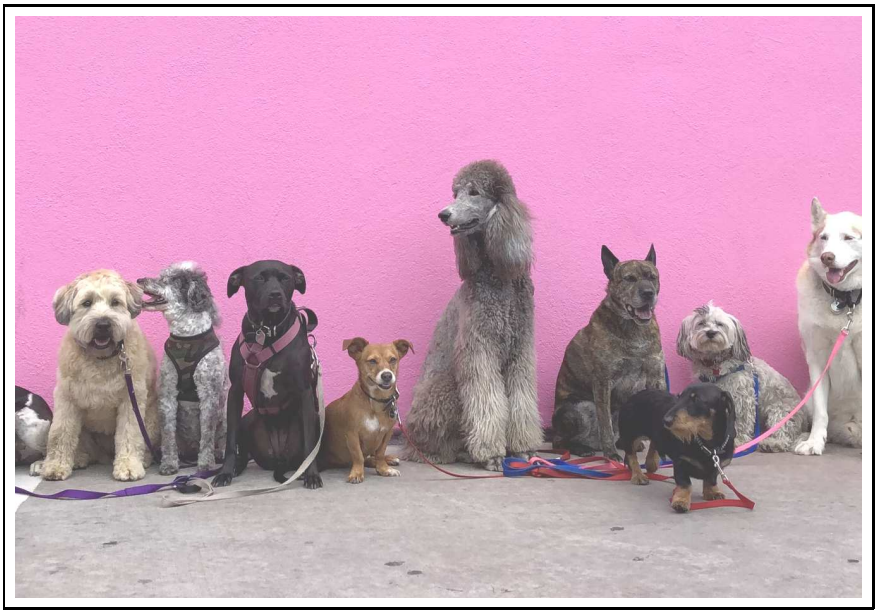
Ciencia de datos energéticos

Descubrimiento de patrones
Agrupamiento

Oscar Duarte, Ph.D.

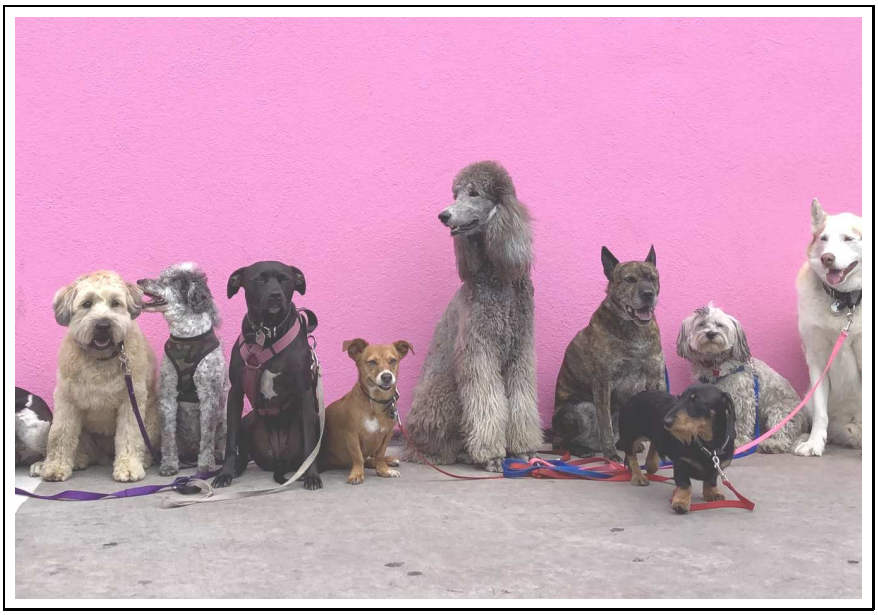
Universidad Nacional de Colombia
Departamento de Ingeniería Eléctrica y Electrónica

Patrones



- ¿Hay perros parecidos?
- ...defina 'parecido'.
- ¿Quién tiene la razón?

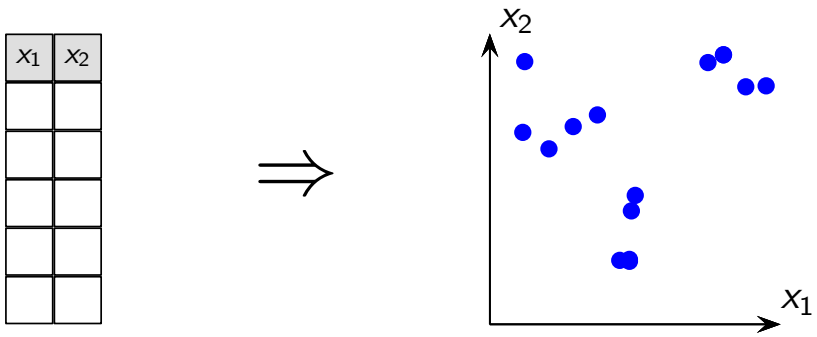
Patrones



- ¿Hay perros parecidos?
- ...defina 'parecido'.
- ¿Quién tiene la razón?

- Patrones.
- Criterio de semejanza.
- Aprendizaje no supervisado.

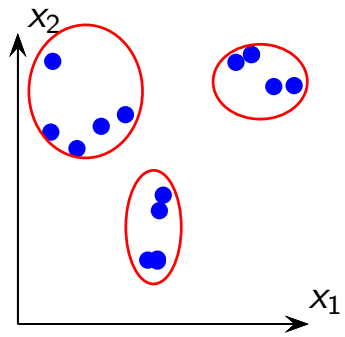
Patrones



Patrones

x_1	x_2

$\Longrightarrow$



Patrones

x_1	x_2	x_3	x_4	x_5	x_6

$\Longrightarrow$

Patrones

x_1	x_2	x_3	x_4	x_5	x_6

$\Longrightarrow$

- Algoritmos de agrupamiento.
- Reglas de asociación.

Algoritmos de agrupamiento
Matriz de distancia

	x_1	x_2	x_3	x_4	x_5	x_6
$X_1 \rightarrow$						
$X_2 \rightarrow$						
$X_n \rightarrow$						

$\Longrightarrow$

Matriz de distancias

	X_1	X_2	$\cdots$	X_n
X_1	$d_{1,1}$	$d_{1,2}$	$\cdots$	$d_{1,n}$
X_2	$d_{2,1}$	$d_{2,2}$	$\cdots$	$d_{2,n}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\cdots$
X_n	$d_{n,1}$	$d_{n,2}$	$\cdots$	$d_{n,n}$

Algoritmos de agrupamiento
Matriz de distancia

	x_1	x_2	x_3	x_4	x_5	x_6
$X_1 \rightarrow$						
$X_2 \rightarrow$						
$X_n \rightarrow$						

$\Longrightarrow$

Matriz de distancias

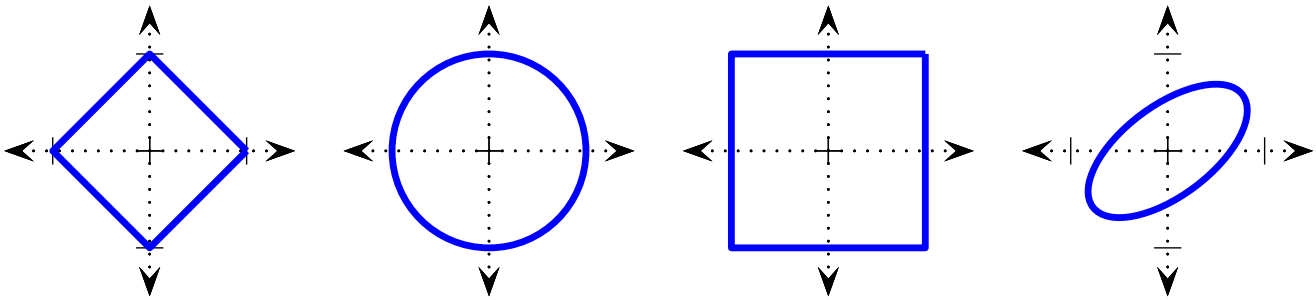
	X_1	X_2	$\cdots$	X_n
X_1	$d_{1,1}$	$d_{1,2}$	$\cdots$	$d_{1,n}$
X_2	$d_{2,1}$	$d_{2,2}$	$\cdots$	$d_{2,n}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\cdots$
X_n	$d_{n,1}$	$d_{n,2}$	$\cdots$	$d_{n,n}$

$$d_{i,j} = \mathcal{D}(X_i, X_j)$$

$$d_{i,j} = \mathcal{N}(X_i - X_j)$$

Algoritmos de agrupamiento
Normas

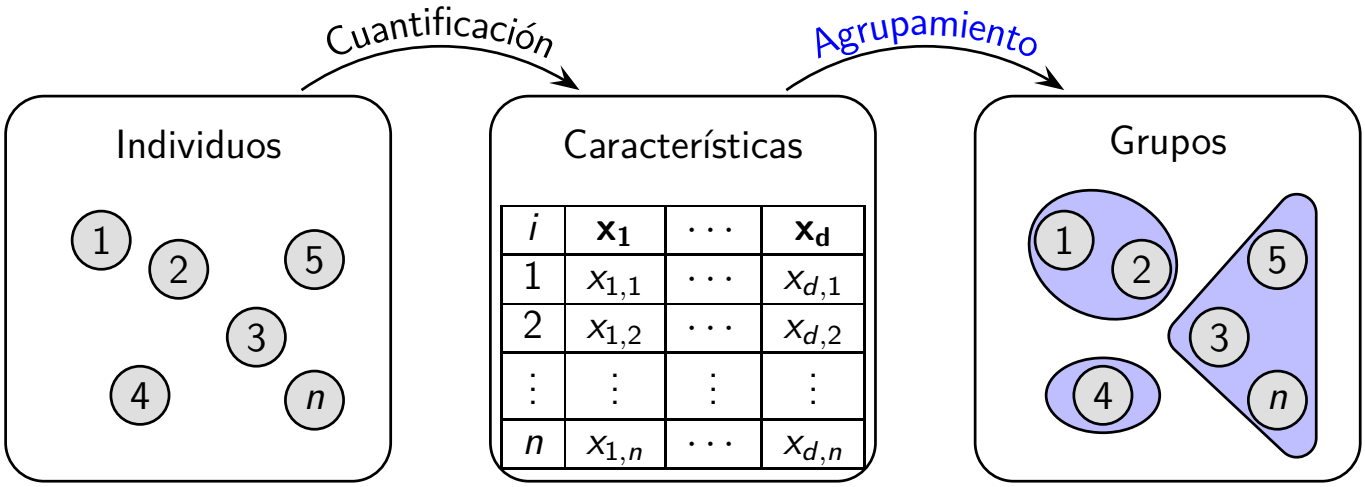
$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{q} x_i^p\right)^{1/p} \qquad q \text{ dimensiones}$$



$\|\cdot\|_1$ $\|\cdot\|_2$ $\|\cdot\|_\infty$ $\|\cdot\|_A$

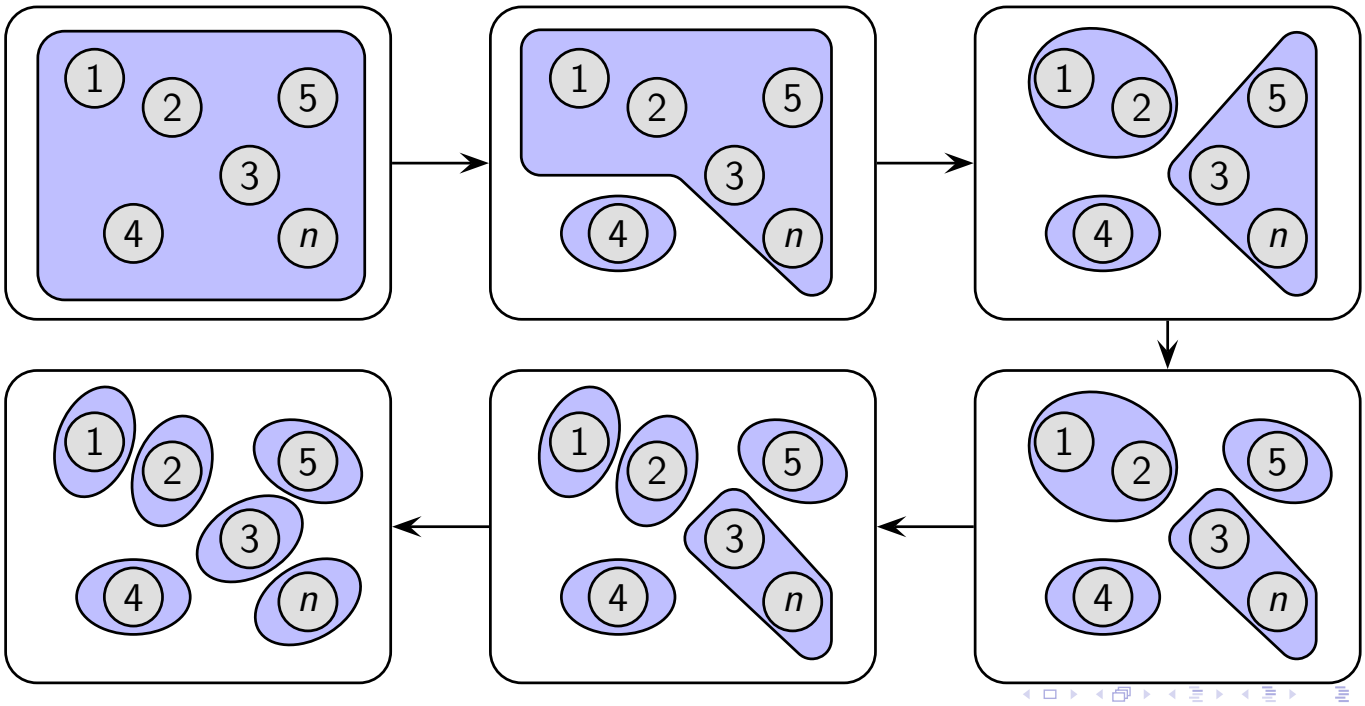
$$A = \begin{bmatrix} 2.08 & -1.44 \\ -1.44 & 2.92 \end{bmatrix}$$

Agrupamiento

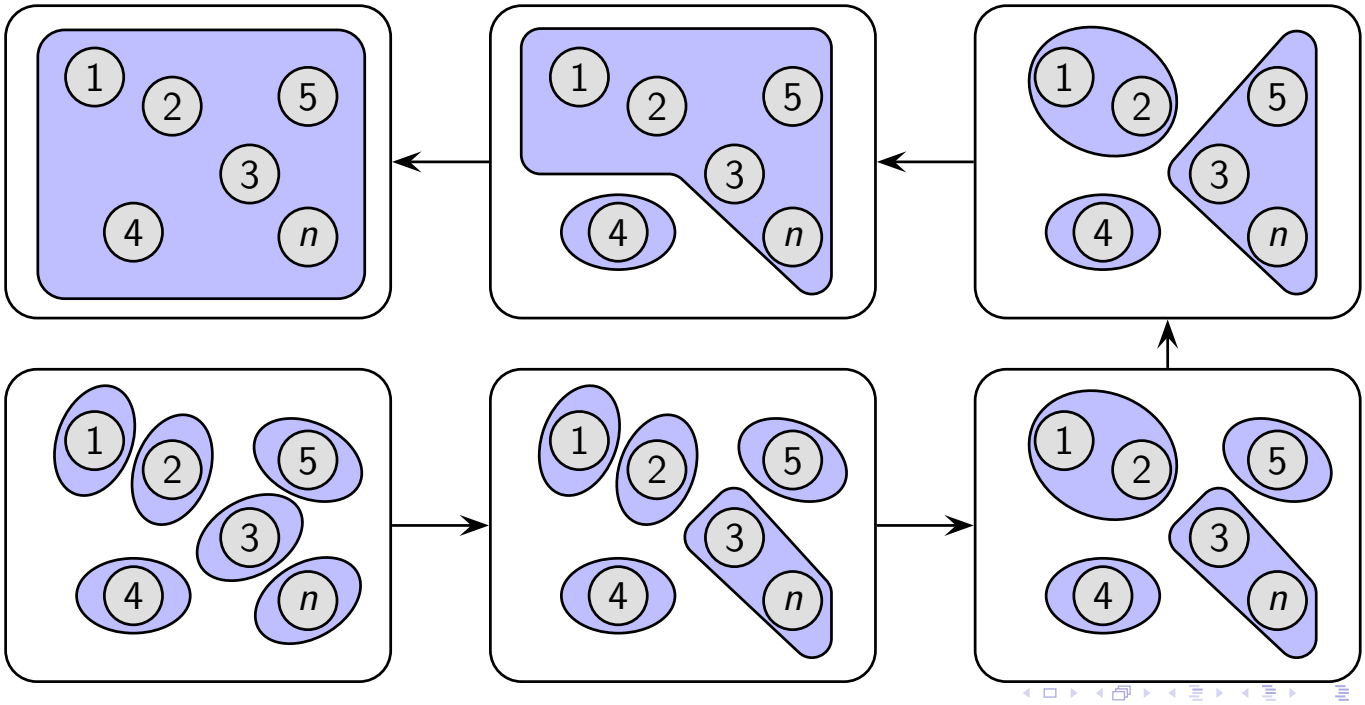


i	$\mathbf{x_1}$	$\cdots$	$\mathbf{x_d}$
1	$x_{1,1}$	$\cdots$	$x_{d,1}$
2	$x_{1,2}$	$\cdots$	$x_{d,2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	$x_{1,n}$	$\cdots$	$x_{d,n}$

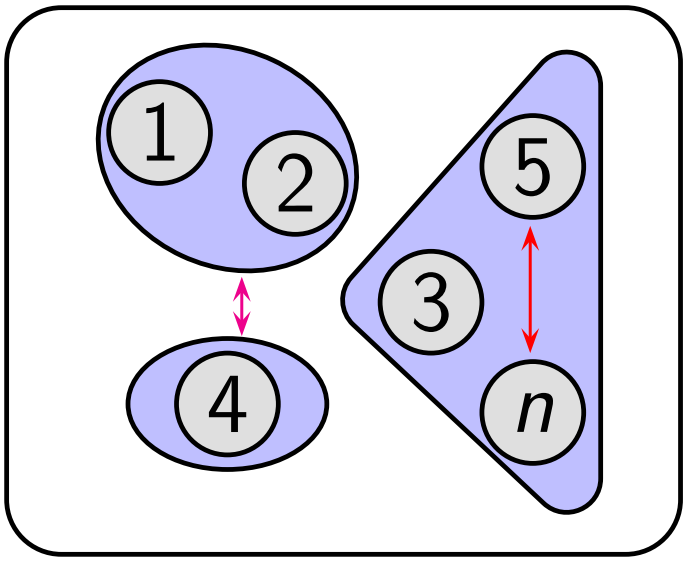
Agrupamiento jerárquico
Divisivo



Agrupamiento jerárquico Aglomerativo



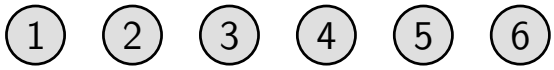
Un buen agrupamiento



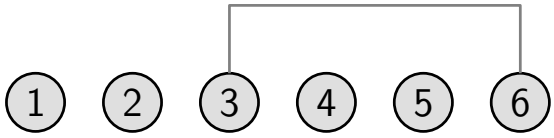
Minimizar distancias intragrupo.

Maximizar distancias intergrupo.

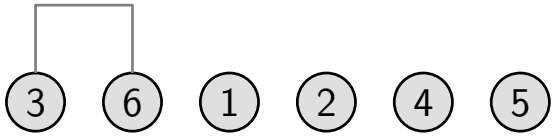
Dendrogramas



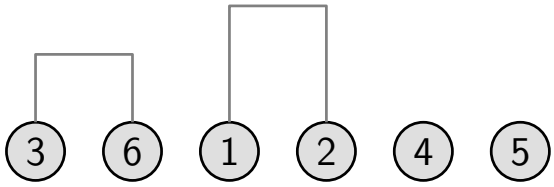
Dendrogramas



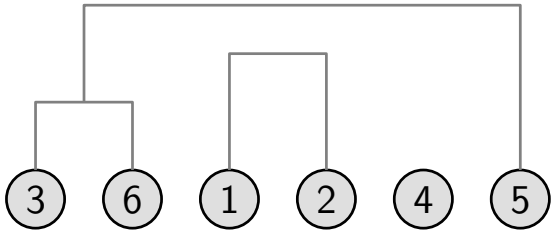
Dendrogramas



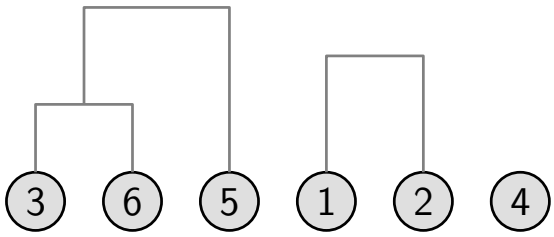
Dendrogramas



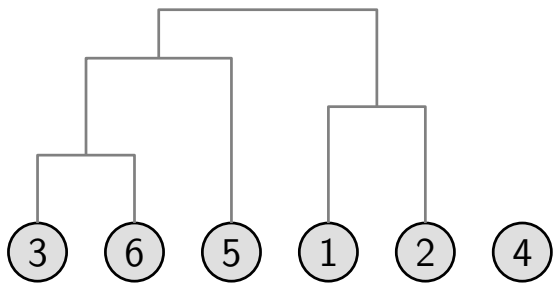
Dendrogramas



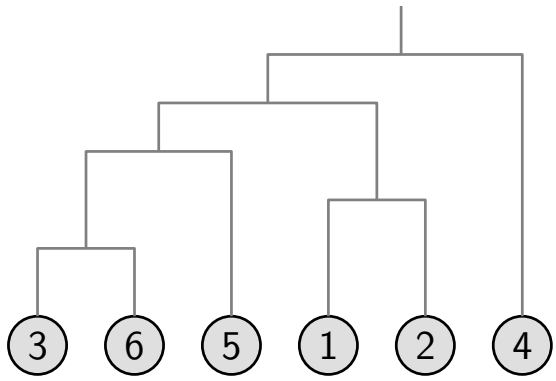
Dendrogramas



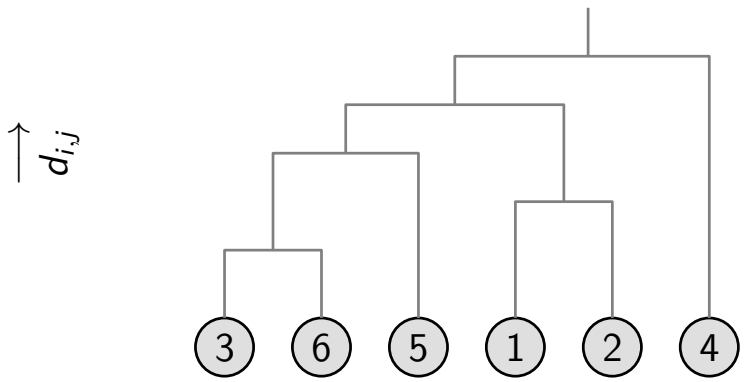
Dendrogramas



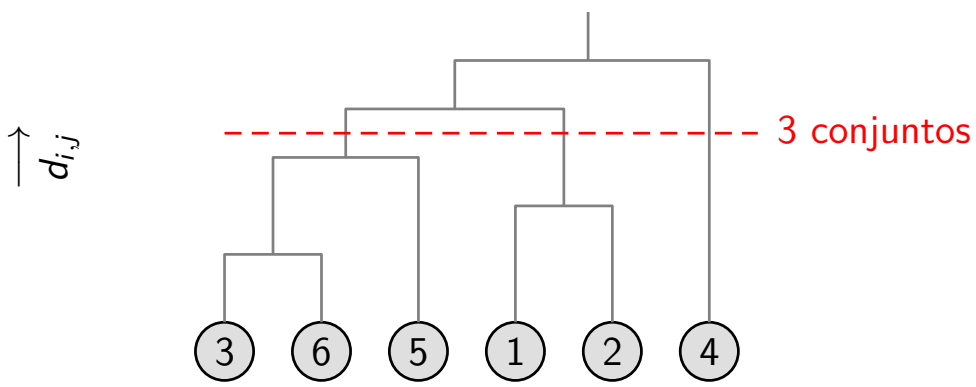
Dendrogramas



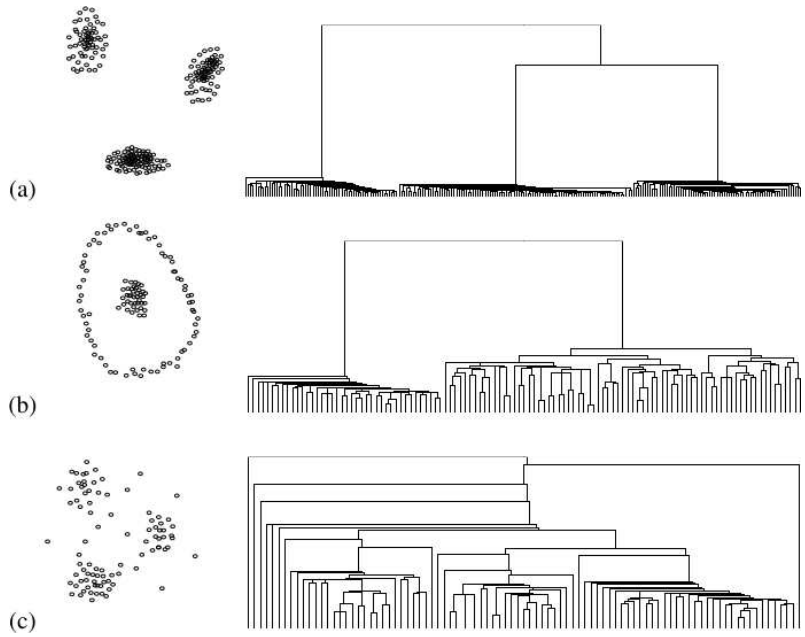
Dendrogramas



Dendrogramas



Dendrogramas



Algoritmo de agrupamiento *k-means*
Construcción de k grupos

x_1	x_2	x_3	x_4	x_5	x_6

$\Rightarrow$

- Cada grupo tendrá un prototipo:

$$\mathbf{p}_i = \frac{\sum p_{i,j}\mathbf{x_j}}{\sum p_{i,j}}$$

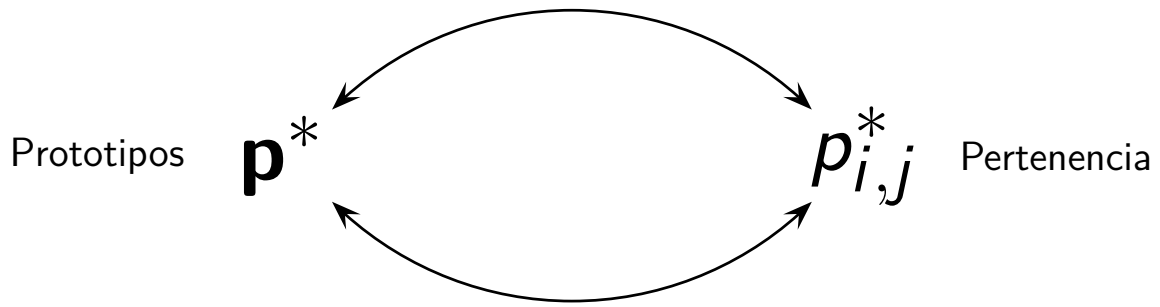
- Cada elemento pertenecerá al grupo del prototipo más cercano:

$$p_{i,j} = \begin{cases} 1 & \text{si } \mathbf{p}_i \text{ es el prototipo más cercano a } \mathbf{x}_j \\ 0 & \text{si otro caso} \end{cases}$$

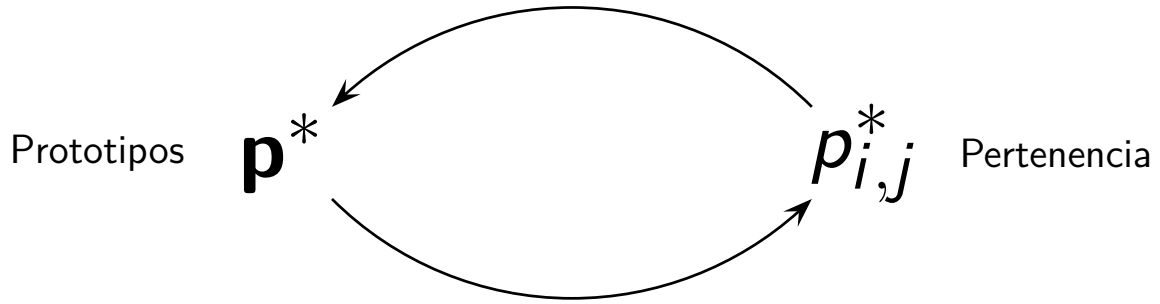
- Minimizar:

$$J = \sum_{i=1}^{k}\sum_{j=1}^{n} p_{i,j}\,\|\mathbf{p}_i - \mathbf{x}_j\|$$

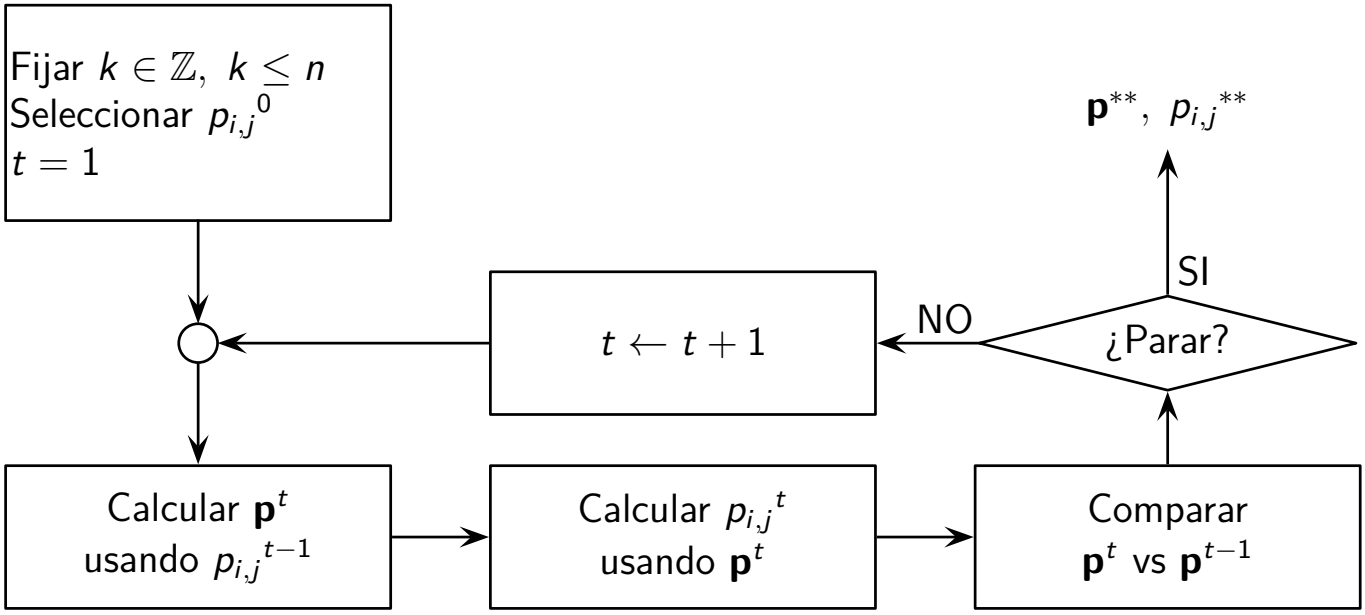
Algoritmo de agrupamiento *k-means*



Algoritmo de agrupamiento *k-means*



Algoritmo de agrupamiento *k-means*
El algoritmo



¡Gracias!

ogduartev@unal.edu.co
@ogduartev

Ciencia de datos energéticos

Descubrimiento de patrones
Reglas de asociación

Oscar Duarte, Ph.D.

Universidad Nacional de Colombia
Departamento de Ingeniería Eléctrica y Electrónica

Reglas de asociación
Principios

Ítems

Transacciones

	x_1	x_2	x_3	x_4
1	1	1	0	1
2	1	0	0	1
3	1	1	1	0
4	1	1	0	1
5	1	1	1	1

Reglas de asociación
Principios

Ítems

Transacciones

	x_1	x_2	x_3	x_4
1	1	1	0	1
2	1	0	0	1
3	1	1	1	0
4	1	1	0	1
5	1	1	1	1

Una regla de asociación tiene la forma:

$$\text{Antecedente} \Rightarrow \text{Consecuente}$$

Antecedente y Consecuente son conjuntos de ítems.Ej:

$$R: \ \{x_1, x_3\} \Rightarrow \{x_2\}$$

Reglas de asociación
Principios

Ítems

Transacciones

	x_1	x_2	x_3	x_4
1	1	1	0	1
2	1	0	0	1
3	1	1	1	0
4	1	1	0	1
5	1	1	1	1

Una regla de asociación tiene la forma:

$$\text{Antecedente} \Rightarrow \text{Consecuente}$$

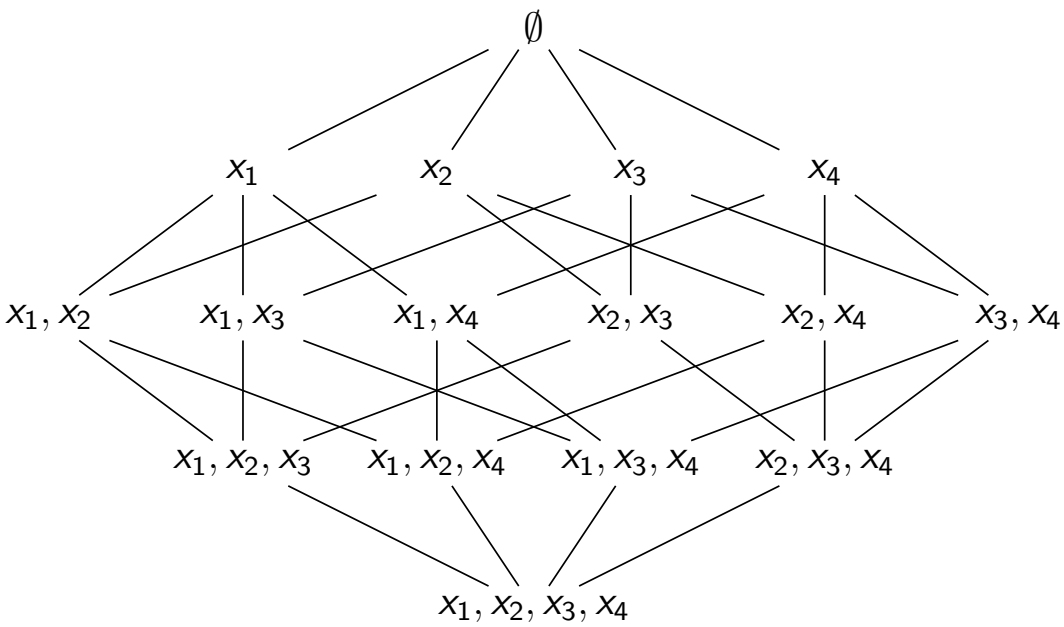
Antecedente y Consecuente son conjuntos de ítems.Ej:

$$R: \ \{x_1, x_3\} \Rightarrow \{x_2\}$$

Interpretación:

En toda transacción que sucede el antecedente, también sucede el consecuente

Retícula de conjuntos de ítems
Item set Laticce



¿Qué reglas son válidas?
Métricas

- Soporte de un itemset:
  sop(I) es el número de transacciones que contienen a I.
- Itemsets frecuentes:
  itemsets cuyo soporte es mayor que un umbral s_{min}.
- Soporte de la regla $R: \ A \rightarrow C$:

$$\text{sop}(R) = \frac{\text{sop}(A \cup C)}{\text{sop}(\emptyset)}$$

- Confianza de la regla $R: \ A \rightarrow C$:

$$\text{conf}(R) = \frac{\text{sop}(A \cup C)}{\text{sop}(A)}$$

Inducción de reglas de asociación

Ítems

Transacciones

	x_1	x_2	x_3	x_4
1	1	1	0	1
2	1	0	0	1
3	1	1	1	0
4	1	1	0	1
5	1	1	1	1

Encontrar el conjunto de reglas de asociación cuyo soporte y confianza excedan unos umbrales específicos:

$$s_{min}, c_{min}$$

Reglas de asociación para tablas no transaccionales

x_1	x_2
2	6
8	4
5	3
3	2
1	6

Reglas de asociación para tablas no transaccionales

x_1	x_2
2	6
8	4
5	3
3	2
1	6

$$x_1 : \begin{cases} x_{1A} : & 0.0 < x_1 \leq 3.3 \\ x_{1B} : & 3.3 < x_1 \leq 6.6 \\ x_{1C} : & 6.6 < x_1 \leq 10 \end{cases}$$

$$x_2 : \begin{cases} x_{2A} : & 0.0 < x_2 \leq 3.3 \\ x_{12} : & 3.3 < x_2 \leq 6.6 \\ x_{2C} : & 6.6 < x_2 \leq 10 \end{cases}$$

Reglas de asociación para tablas no transaccionales

x_1	x_2
2	6
8	4
5	3
3	2
1	6

$$x_1 : \begin{cases} x_{1A} : & 0.0 < x_1 \leq 3.3 \\ x_{1B} : & 3.3 < x_1 \leq 6.6 \\ x_{1C} : & 6.6 < x_1 \leq 10 \end{cases}$$

$$x_2 : \begin{cases} x_{2A} : & 0.0 < x_2 \leq 3.3 \\ x_{12} : & 3.3 < x_2 \leq 6.6 \\ x_{2C} : & 6.6 < x_2 \leq 10 \end{cases}$$

x_1			x_2		
x_{1A}	x_{1B}	x_{1C}	x_{2A}	x_{2B}	x_{2C}
1	0	0	0	1	0
0	0	1	0	1	0
0	1	0	1	0	0
1	0	0	1	0	0
1	0	0	0	1	0

Un ejemplo de aplicación
Análisis de las fallas en circuitos de distribución

Información disponible para una colección de fallas:

- Fecha y hora
- Duración
- Circuito en que ocurrió
- Causa de la falla
- Tipo de falla

Información disponible para cada circuito:

- Ubicación
- Potencia
- Número de transformadores
- Longitud
- Número y tipo de usuarios

Un ejemplo de aplicación
Análisis de las fallas en circuitos de distribución

Información disponible para una colección de fallas:

- Fecha y hora
- Duración
- Circuito en que ocurrió
- Causa de la falla
- Tipo de falla

Información disponible para cada circuito:

- Ubicación
- Potencia
- Número de transformadores
- Longitud
- Número y tipo de usuarios

- *Circuitos cortos tienen fallas que duran poco tiempo*
- *Circuitos con pocos transformadores fallan pocas veces*

¡Gracias!

ogduartev@unal.edu.co
@ogduartev