# Detección de anomalías basada en técnicas de ensembles

Ignacio Aguilera Martos

6 de septiembre de 2019

Trabajo Fin de Grado
Código disponible en GitHub

# Contenidos

# Concepto de anomalía basado en distancias

# Tukey's Fences

# Tukey's Fences

Pensado para el caso uno-dimensional.

# Tukey's Fences

Pensado para el caso uno-dimensional.

**Tukey's Fences**

Valores fuera del rango $[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)]$ con $k = 1{,}5$

# Tukey's Fences

Pensado para el caso uno-dimensional.

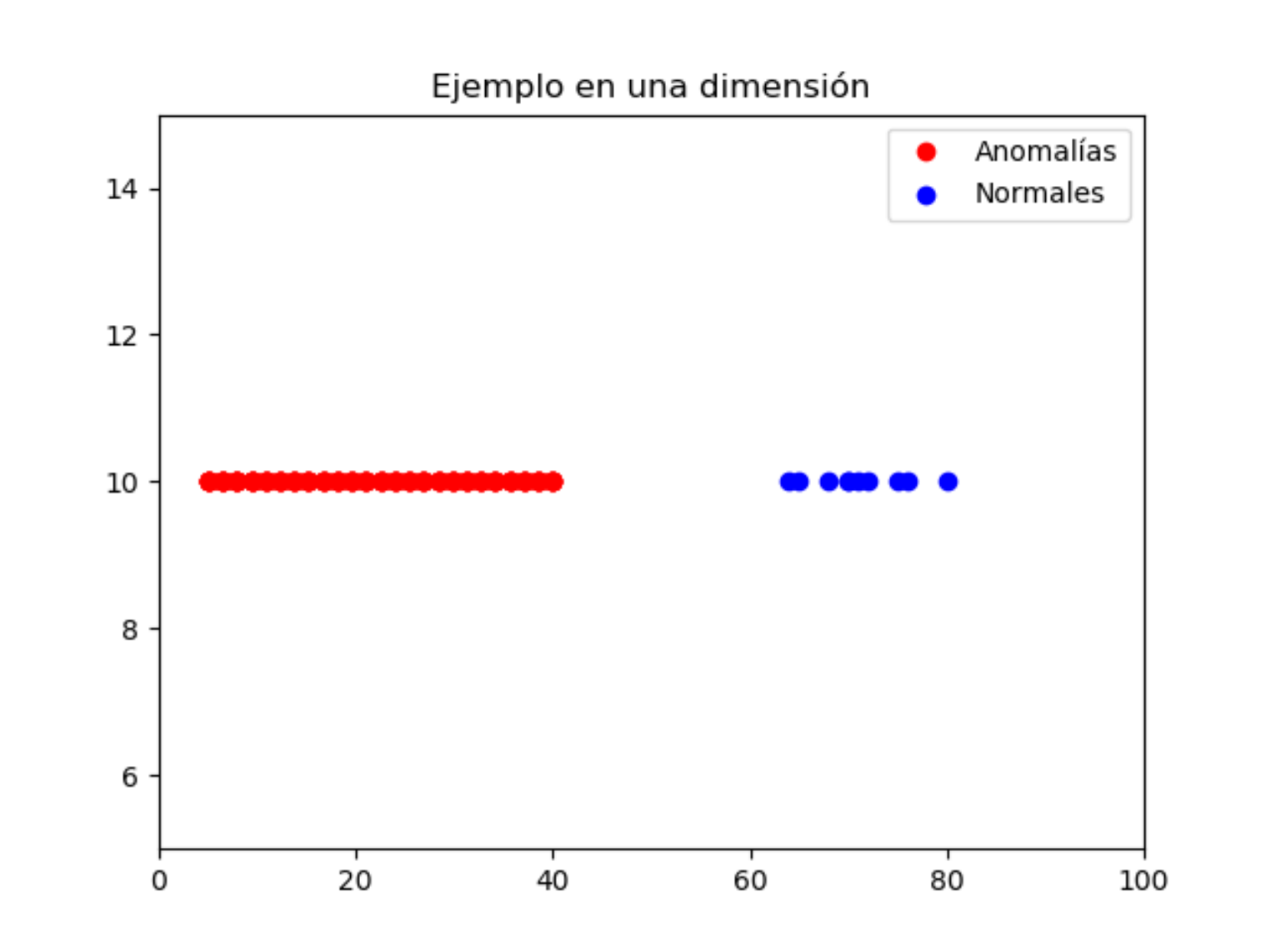
**Tukey's Fences**

Valores fuera del rango $[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)]$ con $k = 1{,}5$

La propuesta de $k = 1{,}5$ es arbitraria.

# Tukey's Fences

# Tukey's Fences



# Extensión al caso de mayor dimensionalidad

# Extensión al caso de mayor dimensionalidad

**Criterio**

Aplicar el criterio de Tukey a cada una de las características.

# Extensión al caso de mayor dimensionalidad

**Criterio**

Aplicar el criterio de Tukey a cada una de las características. **Trivial**.

# Extensión al caso de mayor dimensionalidad

**Criterio**

Aplicar el criterio de Tukey a cada una de las características. **Trivial**.

**Criterio de clusters**

1. Agrupamos los datos por clusters.

# Extensión al caso de mayor dimensionalidad

**Criterio**

Aplicar el criterio de Tukey a cada una de las características. **Trivial**.

**Criterio de clusters**

1. Agrupamos los datos por clusters.
2. Encontramos el cluster más cercano para cada instancia.

# Extensión al caso de mayor dimensionalidad

**Criterio**

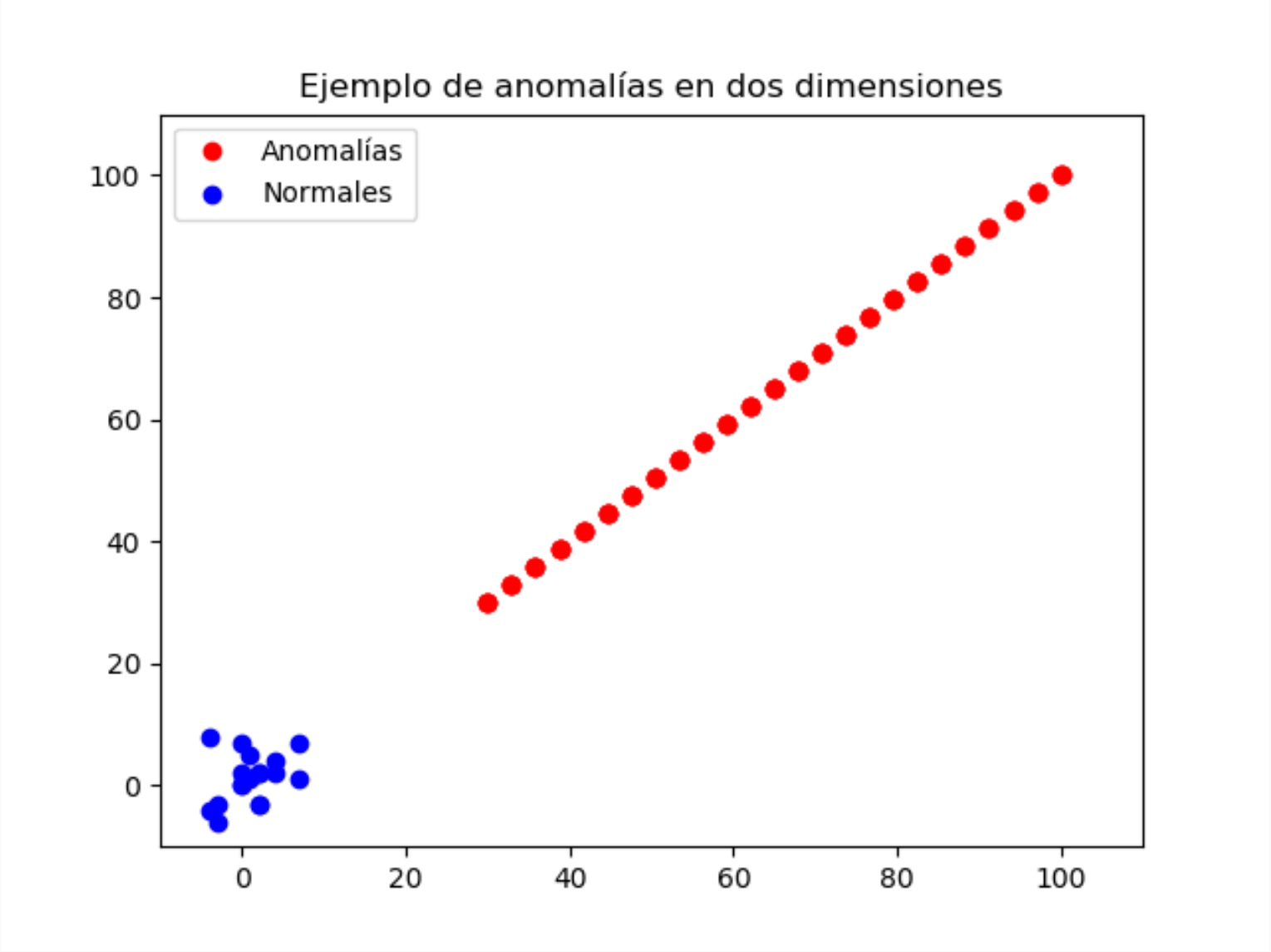
Aplicar el criterio de Tukey a cada una de las características. **Trivial**.

**Criterio de clusters**

1. Agrupamos los datos por clusters.
2. Encontramos el cluster más cercano para cada instancia.
3. Si la distancia del objeto al centroide del cluster es mayor que 1,5 veces la mayor distancia intercluster entonces es una anomalía.

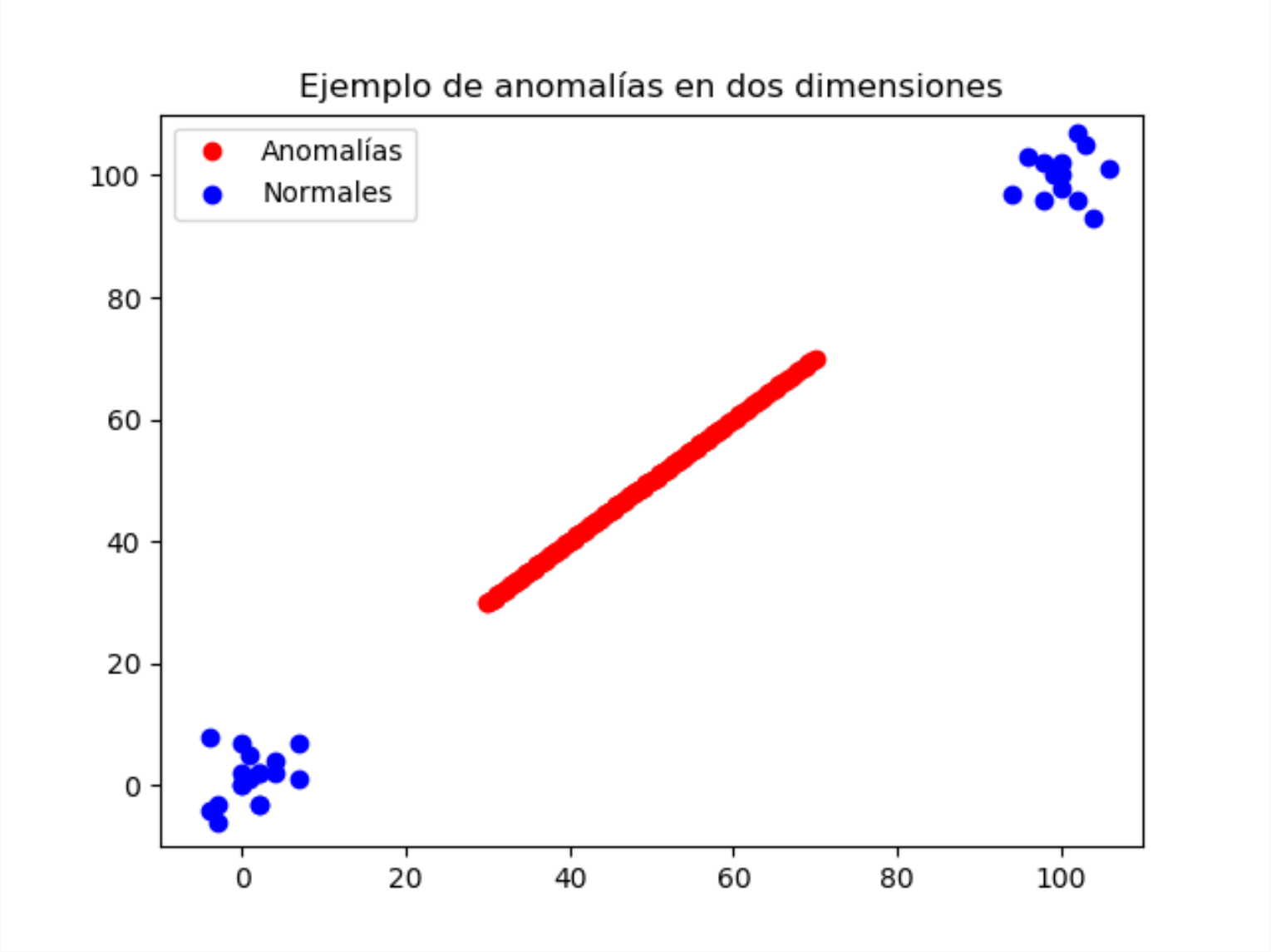
# Ejemplo 1

# Ejemplo 1


Ejemplo de anomalías en dos dimensiones
Anomalías
Normales
100
80
60
40
20
0
0
20
40
60
80
100

# Ejemplo 2

# Ejemplo 2


Ejemplo de anomalías en dos dimensiones
Anomalías
Normales
100
80
60
40
20
0
0
20
40
60
80
100

# Machine Learning

# Problema que abordamos

# Problema que abordamos

**Problema de detección de anomalías**

Es un problema de aprendizaje no supervisado pues no disponemos de las etiquetas.

# Problema que abordamos

**Problema de detección de anomalías**

Es un problema de aprendizaje no supervisado pues no disponemos de las etiquetas.

**Partes teóricas del problema**

1. Generador.
2. Sistema.
3. Máquina de aprendizaje.

# Alta dimensionalidad

# Alta dimensionalidad

## Propiedades de conjuntos de alta dimensionalidad

1. La densidad disminuye exponencialmente al aumentar la dimensionalidad.

# Alta dimensionalidad

## Propiedades de conjuntos de alta dimensionalidad

1. La densidad disminuye exponencialmente al aumentar la dimensionalidad.
2. Cuanto mayor es la dimensionalidad mayor debe ser el radio de una bola para englobar el mismo porcentaje de datos.

# Alta dimensionalidad

## Propiedades de conjuntos de alta dimensionalidad

1. La densidad disminuye exponencialmente al aumentar la dimensionalidad.
2. Cuanto mayor es la dimensionalidad mayor debe ser el radio de una bola para englobar el mismo porcentaje de datos.
3. Casi todo punto está más cerca del borde del conjunto que de otro punto.

$$D(d, n) = (1 - \frac{1}{2}^{\frac{1}{n}})^{\frac{1}{d}}$$

# Alta dimensionalidad

## Propiedades de conjuntos de alta dimensionalidad

1. La densidad disminuye exponencialmente al aumentar la dimensionalidad.
2. Cuanto mayor es la dimensionalidad mayor debe ser el radio de una bola para englobar el mismo porcentaje de datos.
3. Casi todo punto está más cerca del borde del conjunto que de otro punto.

$$D(d, n) = (1 - \frac{1}{2}^{\frac{1}{n}})^{\frac{1}{d}}$$

4. Casi todo punto es una anomalía sobre su propia proyección.

# Alta dimensionalidad

## Propiedades de conjuntos de alta dimensionalidad

1. La densidad disminuye exponencialmente al aumentar la dimensionalidad.
2. Cuanto mayor es la dimensionalidad mayor debe ser el radio de una bola para englobar el mismo porcentaje de datos.
3. Casi todo punto está más cerca del borde del conjunto que de otro punto.

$$D(d, n) = (1 - \frac{1}{2}^{\frac{1}{n}})^{\frac{1}{d}}$$

4. Casi todo punto es una anomalía sobre su propia proyección.

## Maldición de la alta dimensionalidad

A mayor dimensionalidad mayor número de puntos necesitamos para obtener una aproximación con funciones de igual regularidad.

# Aproximación de funciones

# Aproximación de funciones

**Aproximación de funciones**

Nuestro objetivo es aproximar la función de salida del sistema con los datos que tenemos.

# Aproximación de funciones

## Aproximación de funciones

Nuestro objetivo es aproximar la función de salida del sistema con los datos que tenemos.

## Teoremas útiles

- Teorema de Aproximación de Weierstrass

# Aproximación de funciones

## Aproximación de funciones

Nuestro objetivo es aproximar la función de salida del sistema con los datos que tenemos.

## Teoremas útiles

- Teorema de Aproximación de Weierstrass
- Serie de Fourier

# Aproximación de funciones

## Aproximación de funciones

Nuestro objetivo es aproximar la función de salida del sistema con los datos que tenemos.

## Teoremas útiles

- Teorema de Aproximación de Weierstrass
- Serie de Fourier
- Teorema de Kolmogorov-Arnold

# Teorema de Aproximación de Weierstrass

# Teorema de Aproximación de Weierstrass

**Teorema de Aproximación de Weierstrass**

Supongamos que tenemos una función $f : [a, b] \to \mathbb{R}$ continua.
Entonces $\forall \epsilon > 0$ existe un polinomio $p$ tal que $\forall x \in [a, b]$ tenemos que $|f(x) - p(x)| < \epsilon$.

# Teorema de Aproximación de Weierstrass

**Teorema de Aproximación de Weierstrass**

Supongamos que tenemos una función $f : [a, b] \to \mathbb{R}$ continua. Entonces $\forall \epsilon > 0$ existe un polinomio $p$ tal que $\forall x \in [a, b]$ tenemos que $|f(x) - p(x)| < \epsilon$.

**Aproximación por polinomios**

El Teorema de Aproximación de Weierstrass nos da una forma de aproximar funciones por polinomios.

# Serie de Fourier

## Serie de Fourier

Si tenemos una función $f : \mathbb{R} \to \mathbb{R}$ integrable en el intervalo $[t_0 - \frac{T}{2}, t_0 + \frac{T}{2}]$ entonces se puede obtener el desarrollo de Fourier de $f$ en dicho intervalo. Si $f$ es periódica en toda la recta real la aproximación es válida en todos los valores en los que esté definida.

$$f(t) \approx \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[ a_n \cos(\frac{2n\pi}{T} t) + b_n \sin(\frac{2n\pi}{T} t) \right]$$

$$a_0 = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)dt, \; a_n = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) \cos(\frac{2n\pi}{T} t)dt,$$

$$b_n = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) \sin(\frac{2n\pi}{T} t)dt$$

# Teorema de Kolmogorov-Arnold

# Teorema de Kolmogorov-Arnold

**Teorema de Superposición Kolmogorov-Arnold**

Sea $f$ una función continua de varias variables $f : X_1 \times ... \times X_n \to \mathbb{R}$, entonces existen funciones $\Phi_q : \mathbb{R} \to \mathbb{R}$ y $\phi_{q,p} : X_p \to [0, 1]$ tales que $f$ se puede expresar como:

$$f(x) = f(x_1, ..., x_n) = \sum_{q=0}^{2n} \Phi_q \left( \sum_{p=1}^{n} \phi_{q,p}(x_p) \right)$$

# Teorema de Kolmogorov-Arnold

## Teorema de Superposición Kolmogorov-Arnold

Sea $f$ una función continua de varias variables $f : X_1 \times ... \times X_n \to \mathbb{R}$, entonces existen funciones $\Phi_q : \mathbb{R} \to \mathbb{R}$ y $\phi_{q,p} : X_p \to [0, 1]$ tales que $f$ se puede expresar como:

$$f(x) = f(x_1, ..., x_n) = \sum_{q=0}^{2n} \Phi_q \left( \sum_{p=1}^{n} \phi_{q,p}(x_p) \right)$$

## Dimensionalidad en datos y funciones

La maldición de la dimensionalidad es intrínseca a los datos. Este Teorema nos dice que la capacidad expresiva o complejidad de las funciones de una sola variable es la misma que las de varias variables.

# Equilibrio Sesgo-Varianza

# Equilibrio Sesgo-Varianza

## Notación

$$MSE = \frac{1}{n}\sum_{i=1}^{n} y_i - g(X_i, \mathcal{D})^2$$

$$E[MSE] = \frac{1}{n}\sum_{i=1}^{n} E[y_i - g(X_i, \mathcal{D})^2] = \cdots$$

$$= \frac{1}{n}\sum_{i=1}^{n} \{f(X_i) - E[g(X_i, \mathcal{D})]\}^2 + \frac{1}{n}\sum_{i=1}^{n} E[\{E[g(X_i, \mathcal{D})] - g(X_i, \mathcal{D})\}^2]$$

$$= sesgo^2 + varianza$$

# Teoría del Aprendizaje

# Teoría del Aprendizaje

**Teorema clave de la Teoría del Aprendizaje**

Para funciones de pérdida acotadas el principio inductivo de minimización del error empírico es consistente sí y sólo si el error empírico converge uniformemente al valor real del error en el siguiente sentido:

$$\lim_{n\to\infty} P[\sup_{\omega} |R(\omega) - R_{emp}(\omega)| > \epsilon] = 0 \ , \ \forall \epsilon > 0$$

# Cota ERM

# Cota ERM

**Dimensión VC**

Decimos que un conjunto de funciones tiene dimensión VC $h$ si puede resolver de forma óptima todos los casos de tamaño $h$ pero existe al menos uno de tamaño $h + 1$ que no puede resolver.

# Cota ERM

## Dimensión VC

Decimos que un conjunto de funciones tiene dimensión VC $h$ si puede resolver de forma óptima todos los casos de tamaño $h$ pero existe al menos uno de tamaño $h+1$ que no puede resolver.

## Cota ERM

Con probabilidad $1-\eta$

$$R(\omega) \leq R_{emp}(\omega) + \frac{\epsilon}{2}\left(1+\sqrt{1+\frac{4\cdot R_{emp}(\omega)}{\epsilon}}\right)$$

# Cota ERM

## Dimensión VC

Decimos que un conjunto de funciones tiene dimensión VC $h$ si puede resolver de forma óptima todos los casos de tamaño $h$ pero existe al menos uno de tamaño $h+1$ que no puede resolver.

## Cota ERM

Con probabilidad $1-\eta$

$$R(\omega) \leq R_{emp}(\omega) + \frac{\epsilon}{2}\left(1+\sqrt{1+\frac{4\cdot R_{emp}(\omega)}{\epsilon}}\right)$$

$$\epsilon = a_1 \cdot \frac{h(\ln(\frac{a_2 n}{h})+1) - \ln(\frac{\eta}{4})}{n}$$

# ¿Preguntas?