**Clase 1 A**

**IA**

🡪 Inteligencia llevada a cabo por maquinas

**Deductiva** 🡪 Lógica 🡪 Sistemas expertos

**Inductiva** 🡪 Ejemplos 🡪 Redes neuronales 🡪 Aprendizaje automático

Aprendizaje automático 🡪 técnicas que brindan a las computadoras la capacidad de aprender **sin haber sido programadas** explicitamente

Diagrama

Descripción generada automáticamente

**Aprendizaje supervisado**

🡪 Respuesta discreta/continua

🡪 Problema de clasificación/regresión

🡪 **Predicción de un resultado futuro** a partir de datos disponibles

Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamente

🡪 Respuesta **discreta** 🡪 Problema de **clasificación**

🡪 Respuesta **continua** 🡪 Problema de **regresión**

**Redes neuronales**

🡪 Buscan emular el comportamiento del cerebro humano

🡪 Neuronas interconectadas

Diagrama

Descripción generada automáticamenteEntradas Xi 🡪 señales de otras neuronas

Pesos Wi 🡪 Intensidad de la sinapsis que conecta 2 neuronas

Θ 🡪 Función umbral que la neurona debe sobrepasar para activarse

**Tipos de datos disponibles**

**Tipos de variables**

Cuantitativas 🡪 Discretas/continuas

Cualitativas 🡪 Nominales/Ordinales

**Descripciones estadísticas básicas**

**Media** 🡪 Promedio de los valores del atributo 🡪 Numérico

**Mediana** 🡪 **Divide a los valores del atributo en dos partes iguales** de manera que los anteriores son todos menores que el y los siguientes son mayores

**Moda** 🡪 Valor que aparece con mayor frecuencia 🡪 Puede haber mas de 1

**Rango medio** 🡪 Media de los valores max y min del conjunto

**Medidas de dispersión**

**Varianza** 🡪 Dispersion de datos con respecto a la media

🡪 Valores bajos 🡪 Cerca de la media

🡪 Valores altos 🡪 Muy dispersos

**Desviacion estandar** 🡪 Raiz cuadrada de la varianza

**Rango** 🡪 Diferencia entre valores max y min del conjunto

**Cuantiles – Cuartiles – Percentiles**

**Cuantiles** 🡪 Valores que dividen un conj. numérico ordenado en partes iguales

**Cuartiles** 🡪 Dividen la distribución en 4

**Deciles** 🡪 Dividen la distribución en 10

**Centiles** 🡪 Dividen en 100

Diagrama

Descripción generada automáticamente con confianza media

Principio del formulario

**Rango intercuartil** 🡪 Distancia entre Q1 y Q3 🡪 **Rango cubierto por la mitad de los datos**

**Gráficos**

* Diagrama de barras
* Grafico torta

**Histograma** 🡪 Indican la frecuencia de una variable 🡪 Variables cuantitativas

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente**Diagrama de caja**

**Valores atípicos**

**Leves 🡪** [Q1 – 3\*RIC ; Q1 – 1.5\*RIC`] o (Q3 + 1.5\*RIC ; Q3 + 3\*RIC]

**Extremos** 🡪 Inferiores a Q1 – 3\*RIC / Superiores a Q3 + 3\*RIC

**Diagrama de dispersión**

🡪 Pares de valores en un eje coordenado

🡪 Indica la **relación lineal entre X e Y**

**Coeficiente de correlación lineal** 🡪 Valor que indica si 2 atributos están linealmente relacionados o no

**Covarianza** 🡪 Indica el grado de variación conjunta de dos v.a respecto a sus medias

**Clase 1B – Preprocesamiento de datos**

**Preparación de datos**

* Datos faltante
* Valores extremos
* Inconsistencias
* Ruido

**🡪 Limpieza**

**🡪 Transformación**

**Limpieza de los datos**

🡪 Distintos tipos

🡪 Análisis para cada tipo

**Variables con ruido** 🡪 Valores que caen fuera del rango 🡪 outliers

**Valores faltantes** 🡪 Se ignoran/rellenan manualmente/Usar cte. Global/Usar valor medio

**Transformación de atributos**

**Discretizacion**

🡪 Convierte atributos numéricos en ordinales

Ejemplo: Aprobado [4, 10], Desaprobado [0, 4]

**Por rango**

**🡪** Dividir el rango del atributo en k partes iguales

🡪 Cada parte = (Max – Min)/k

**Por frecuencia**

🡪 Dividir los valores **en k partes con la misma cantidad de valores** en cada una

🡪 Atributo debe tener k valores diferentes

**Numerizacion**

🡪 Contrario a discretización 🡪 Cualitativos en numéricos

🡪 Para nominales se usa representación binaria 🡪 Para ordinales entera

**Numerizacion binaria (dummy)**

🡪 Reemplaza al atributo **nominal por atributos binarios como valores distintos pueda tomar**

**🡪** Se iguala el nombre original con alguno de los posibles valores

**Normalizacion**

🡪 Expresa los valores de los atributos sin utilizar las unidades originales

**Lineal uniforme**

🡪 Sensible a outliers

🡪 Pueden obtenerse valores negativos y/o mayores a 1 🡪 Si se recortan los extremos

**Estándar (media y desvio)**

🡪 Tipificacion de datos 🡪 La nueva variable tiene media 0

🡪 Se distribuyen **normalmente alrededor de 0 y con desviación 1**