# ­­­­APRENDIZAJE AUTOMATICO

## TEMA 2 – INTRODUCCION

* ¿Qué es el aprendizaje automático? ✅
  + Nuevos conocimientos y habilidades
  + Mejora continuamente
* Motivación✅
  + Manejar mucha cantidad de datos
  + Capacidad de procesamiento o computación
* Tipos de problemas✅
  + Clasificación
  + Regresión / Predicción
  + Clustering / Agrupamiento
  + Recomendación
  + Recuperación de la información
  + Elaboración de estrategias
  + Declaración de anomalías
* Aplicaciones✅
  + Descubrimientos científicos
  + Medicina
  + Personalización / Recomendación
  + Minería de datos
  + Generación artística
  + Visión artificial
* Tipos AA✅
  + Supervisado - se conoce la salida (predecir)
  + No supervisado – explorativo no se sabe la salida
  + Refuerzo – modelo iterativo reforzar o penalizar tendencias
* Cuestiones éticas✅

## TEMA 3 – FUNDAMENTOS

* Atributo (columnas), instancia (filas), clase (salida/variable dependiente), ejemplo (mismo que instancia) ✅
* Generalización (objetivo), hipótesis (modelo) ✅
* Razonamiento✅
  + Deductivo: Partimos de los general a lo concreto
  + Inductivo: Partimos de lo concreto a lo general
  + Hipotético-deductivo
  + Abductivo
  + Transductivo: Partimos de los concreto a lo concreto
* Ampliaciones✅
  + Emsemble learning 🡪 Grupos de modelos – Votación
  + Active learning 🡪 Operador humano que controla el aprendizaje
  + Online learning 🡪 Flujo continuo de datos
  + Transfer learning 🡪 Partir de un modelo y reentrenarlo
* Pipeline 🡪 Proceso cíclico✅
  + 1. Datos
  + 2. Entrenamiento
  + 3. Evaluación
* Datos (Preprocesamiento de datos) ✅
  + Obtención, fusión, transformación y filtración.
  + ¿Datos imperfectos? (outliers, unknow, ruido)
  + ¿Transformación? (normalización, estandarización, binarización, selección de instancias/atributos)
  + Balanceo (undersampling/descartar, oversampling/generar, SMOTE/heurística al generar)
  + Fuga de datos
    - Uso de información externa (info futura)
    - Bajo: Dummy, eliminación atributos con varianza 0 y eliminación de atributos identificativos
    - Medio: Imputación o escalado
    - Alto: Selección de atributos basados en modelos
* Evaluación
  + Estimar el comportamiento a futuro
  + Evaluación cruzada
    - Leave p out
      * Leave 1 out
    - Holdout
    - K-fold
      * Stratified k-fold
      * Montecarlo
    - Time-series
    - Neasted
  + Matriz de confusión (combinación de casos posibles)
    - Verdaderos positivos (Si 🡪Si)
    - Falsos negativos (No 🡪Si)
    - Falsos positivos (Si 🡪No)
    - Verdaderos negativos (No 🡪No)
  + F1-Score
    - Sensibilidad / Exactitud
  + Curva ROC
    - Sensibilidad / Especificidad
    - AUC (malo cuando se acerca a 0.5)
  + Medidas de error: MSE, MPSE, MAE, …
  + Tipos de error (minimizar sesgo y varianza a la vez):
    - Irreducible
    - Sesgo
    - Varianza
  + Ajustes de hiperparámetros

## TEMA 4 – ÁRBOLES DEDECISION

* Supervisado✅
* Clasificación, Regresión✅
* CLS, ID3, CART, C4.5, CTree, Random Forest✅
* IDE3 (En qué orden se seleccionan los atributos) ✅
  + Ganancia de información
  + Heurística 🡪 Entropía (Desorden de los datos)
  + InfGain = H(nodo padre) – sum(n\_i/n H(nodo\_hijo\_i))
  + H(atributo) 🡪 H(Valores)
  + GainRatio = InfGain/ Entropía
  + Gini más fácil de calcular que GainRatio
* Fronteras de decisión (plano que separa las clases) ✅
* Poda✅
  + Pre: Durante Entrenamiento
  + Post: Después
* Atributos multievaluados 🡪Binarización
* Atributos numéricos 🡪Rangos 🡪Binarización / Multievaluados
* Coste de los artibutos
* Hiper parámetros
* CART, C4.5
* **TABLA FINAL**

Tabla

Descripción generada automáticamente

## TEMA 5 – REGRESIÓN

* Supervisado
* Predecir (hiperplano define )
* y(x) = f(x) + E
  + Ajustar una curva a los puntos
* Fuentes de ruido
  + Imprecisión en los valores de los atributos de entrada
  + Imprecisión en los valores de los atributos de salida
  + Atributos latentes: atributos que no están
  + Modelo simple
* Descenso del gradiente
  + Actualizar los pesos w en función del error
  + Estocástica por lotes
* Regresión lineal
* Regresión polinómica – Orden M (mayor M mayor sobreajuste)
  + Regularización – ajustar el peso de lso coeficientes
    - Regularización Ridge (reducir la varianza)
    - Regularización Lasso (descartar coeficientes menos influyentes)
    - Elastic Net, factor alpha = 0 (Ridge) alpha = 1 (Lasso)
* Distribución de datos 🡪Normal, Poisson, Binomial, Exponencial…
* Regresión logística (enfoque probabilístico)

## TEMA 6 – SVM

* Supervisado
* Clasificación
* Buscar el hiperplano que separa las clases maximizando el margen entre ellas
  + Vectores de soporte (instancias cercanas a la clase contraria)
  + Maximizar el margen
  + Minimizar la inversa del margen al cuadrado
  + Lo que está contenido en el margen es indefinido
* Variables de holgura
  + Hiperparámetro C
* Aumentar la dimensionalidad
  + Problema de computación
    - Kernel trick
    - Kernels: lineal, polinómico, gaussiano
* ¿Multiclase?
  + One vs All
  + One vs One

## TEMA 7 – IBL

* Supervisado
* No crea un modelo 🡪 Lazy learning
* KNN
  + Similitud, distancia
  + 1NN: Mapas de voronoi
    - Rangos de las variables
    - Métrica de la distancia
    - Variables
  + Si hay muchas instancias…
    - Indexación
    - Selección de instancias (Edición Wilson)
    - Mejores ejemplos (condensación)