

# Tarea: Artículo Statical Review Pattern Recognition

Alumno: Alfonso Murrieta Villegas

Semestre; 2022 -1

Grupo: 1

## 1. ¿Qué es reconocimiento de patrones?

De manera general es procedimiento de las ciencias de la computación y otras áreas que se encarga de estudiar y proponer métodos y algoritmos para extraer información de un *dataset* o conjunto de datos independientemente del formato o tipo que sean.

## 2. De acuerdo a la Tabla 2 cuales son los Modelos principales de reconocimiento de patrones. De una explicación de 2 líneas de cada uno.

### 1. *Template Matching*

Es un método empleando para determinar que tan parecidos son 2 entidades, donde una entidad es conocida para ser comparada con la otra, es decir, determina lo similar que son ambas entidades para así determinar la correlación de estas.

### 2. *Syntactic or structural approach*

Al igual que divide y vencerás, es un método donde se trata un patrón complejo en varios sub-patrones o también denominados primitivas.

Los patrones son visto como oraciones pertenecientes a un lenguaje y las primitivas como las letras del alfabeto del mismo lenguaje, de esta forma structural approach trata de explicar este patrón en relación de las primitivas.

### 3. *Statistical Approach*

Es un método donde cada patrón se aborda mediante características o medidas y su efectividad depende directamente de que tan bien se puedan separar los patrones de cada clase a analizar

### 4. *Neuronal networks*

Por último, las redes neuronales de manera general es una entidad que busca a través de nodos encontrar la relación entre las variables de entrada en la fase de entrenamiento, que además se caracteriza por un proceso que trabaja en paralelo.

Cabe destacar que este método se caracteriza principalmente por abordar relaciones que no son lineales

### 3. De la figura 1 de una breve explicación del diagrama

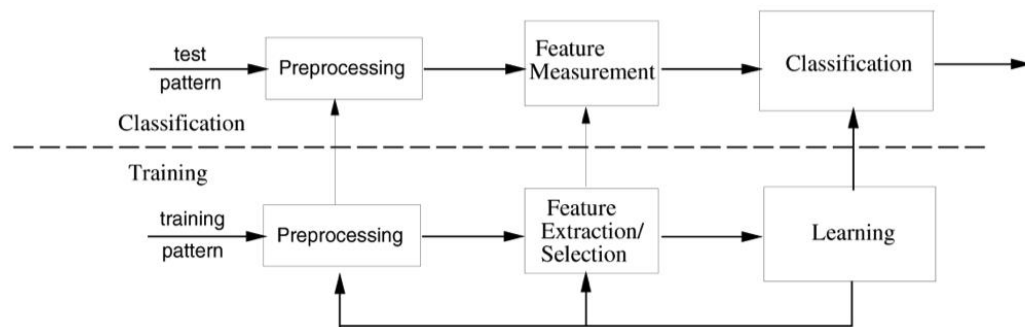


Fig. 1. Model for statistical pattern recognition.

Lo que observamos es que de nuestro dataset o conjunto de datos debemos considerar un split o división de los datos para de esa forma tener un subconjunto para la fase o etapa de entrenamiento y otro para la etapa de prueba.

Posterior a ello, el conjunto de entrenamiento pasa por una etapa de pre-procesamiento donde principalmente se hace merge-data y limpieza de datos, para posteriormente hacer la selección de variables con las que se hará el modelo que entrará en una fase de aprendizaje.

Por otro último, en la etapa de clasificación lo que se hace es probar a través del conjunto de prueba o test que igual previamente fue preprocesado.

### 4. ¿Qué es sobreajuste y subajuste?

Ambos son consecuencias de un mal ajuste en el entrenamiento de un modelo, en el caso de un sobreajuste u overfitting es cuando el modelo está ajustado para un conjunto concreto, es decir, que se entrenó muy concretamente para los datos de entrenamiento, por otro lado, en el caso de un subajuste o underfitting sucede cuando un modelo es demasiado simple, lo que puede ser el resultado de un modelo que necesita más tiempo de entrenamiento, más funciones de entrada o menos regularización.

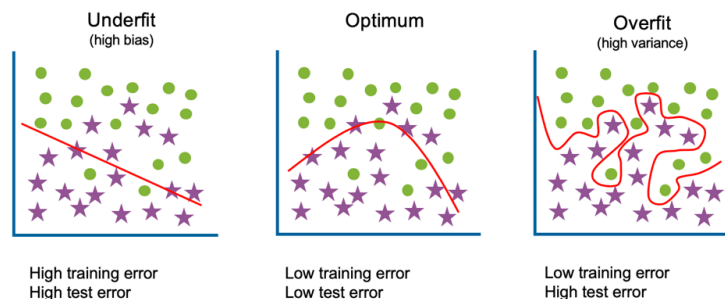


Imagen 1: tipos de casos en el entrenamiento de modelos

## **5. ¿Qué tipo de algoritmos se utilizan para el análisis exploratorio de datos?**

Principalmente:

1. Algoritmos de clustering
  - a. Enfocado principalmente en la clasificación o categorización de conjuntos
2. Plotting o visualización de datos
  - a. Visualización de la distribución de los datos
3. Análisis de datos multidimensionales.
  - a. Principalmente para la reducción de dimensiones para su posterior tratamiento

## **6. ¿Qué es el conjunto de datos Iris ?**

Es un dataset o conjunto de datos que contiene 3 tipos de variantes de Iris (Setosa, virginica y versicolor) a través de un muestreo de 50 casos por cada variante o clase.

Además, este dataset se caracteriza por contener 4 principales atributos, el largo y ancho del sépalo y pétalo

## **7. Diga que entiende por la maldición de la dimensionalidad.**

También denominado como “Fenómeno de aparición” es una paradoja donde teóricamente la probabilidad de que una regla de decisión clasifique de forma errónea no incrementa cuando el número de características incrementa también, sin embargo, en la realidad entre más características hay el rendimiento del modelo empeora

## **Referencias**

[1] Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: A review. IEEE Transactions on pattern analysis and machine intelligence, 22(1), 4-37.

[2] What is underfitting?. IBM. Recuperado 28 de septiembre de 2021, de <https://www.ibm.com/cloud/learn/underfitting>