

Reducción de la Dimensionalidad

Minería de Datos

Máster en Tecnologías de Análisis de Datos Masivos

José T. Palma

Departamento de Ingeniería de la Información y las Comunicaciones
Universidad de Murcia

2025



UNIVERSIDAD
DE MURCIA



Contenidos de la presentación

- 1 Introducción
- 2 Selección de Características
 - Fases
 - Generación de subconjuntos
 - Evaluación
- 3 Evaluadores de características: “Rankers”
- 4 Métodos basados en filtros
 - Medidas de relevancia
- 5 Métodos basados en envoltura (“wrappers”)
- 6 Métodos embebidos
- 7 Conclusiones

Introducción

- Como ya es conocido en la actualidad, los datos disponibles para análisis se van acumulando a una velocidad sin precedentes.
- Además, la minería de datos no tiene que ser el único objetivo de la recogida de datos.
- Por lo tanto, el preprocesamiento, y en especial la reducción de la dimensionalidad es un aspecto crucial para una minería de datos eficiente.
- Concretamente, la selección de características (atributos o variables) es un aspecto crucial para reducir el tamaño de los datos.

Introducción: ¿Por qué reducir la dimensionalidad? I

- La complejidad de la mayoría de algoritmos de aprendizaje depende de la dimensión del conjunto de entrada (número de variables) y del número de instancias.
 - **La maldición de la dimensionalidad:** A medida que aumentamos el número de variables se reduce la densidad de los datos, con lo que la determinación de los hipersuperficies de clasificación se vuelve más difícil.
 - La precisión de las consultas y su eficiencia se degrada rápidamente a medida que aumenta la dimensionalidad.
 - La representatividad de los datos (su distribución) requiere de un número de instancias que crece exponencialmente con su dimensión.
 - Modelos simples son más robustos en conjuntos pequeños.

Introducción: ¿Por qué reducir la dimensionalidad? II

- La dimensión intrínseca del problema suele ser pequeña.
 - Sólo un subconjunto de las variables de entrada suele tener relevancia para el proceso de aprendizaje.
- Cuando tenemos menos variables para explicar los datos:
 - nos podemos hacer una mejor idea de los procesos que generan dichos datos, y
 - facilitamos la extracción de conocimiento.
- Al representar los datos en unas pocas dimensiones la representación gráfica y su análisis visual es más sencillo.

Métodos I

- Existen dos tipos de métodos para la reducción de la dimensionalidad: *Extracción de Características* y *Selección de Características*.
- **Extracción de Características.**
 - **Objetivo:** Encontrar k dimensiones que sean combinación de las d ($d > k$) dimensiones originales.
 - Básicamente consiste en encontrar una transformación desde un espacio de dimensión d a un espacio con menos dimensiones, k .
 - Hacen uso de todas las dimensiones originales.
 - Supervisados: maximizar la discriminación entre clases.
 - No supervisados: minimizar la pérdida de información.

Métodos II

- Técnicas:
 - Análisis lineal discriminante (LDA)
 - Análisis discriminante generalizado (GDA)
 - Análisis de componentes principales (PCA) y su versión con funciones Kernel.
 - Modelos de variables latentes basados en procesos gaussianos constreñidos
 - “t-distributed stochastic neighbor embedding” (t-SNE)
 - “Uniform Manifold Approximation and Projection” (UMAP)
 - Técnicas basadas en Deep Learning: Convolución, embeddings, transformers,...

Métodos III

- **Selección de Características.**

- **Objetivo:** Encontrar que k dimensiones, de las d dimensiones originales, aportan la mayor cantidad de información, descartando el resto ($d - k$).
- Básicamente, se trata de encontrar el subconjunto óptimo de dimensiones de acuerdo con alguna determinada función objetivo.
- Sólo se selecciona un subconjunto de las dimensiones originales.
- Se trabaja directamente con las dimensiones originales sin realizar transformación alguna.

Selección de Características

- La **Selección de Características** trata de encontrar el subconjunto de características óptimo de acuerdo con un determinado criterio de selección.
- Dicho subconjunto óptimo de características debe garantizar que el proceso de construcción de modelos se pueda desarrollar con totales garantías, minimizando el número de características.

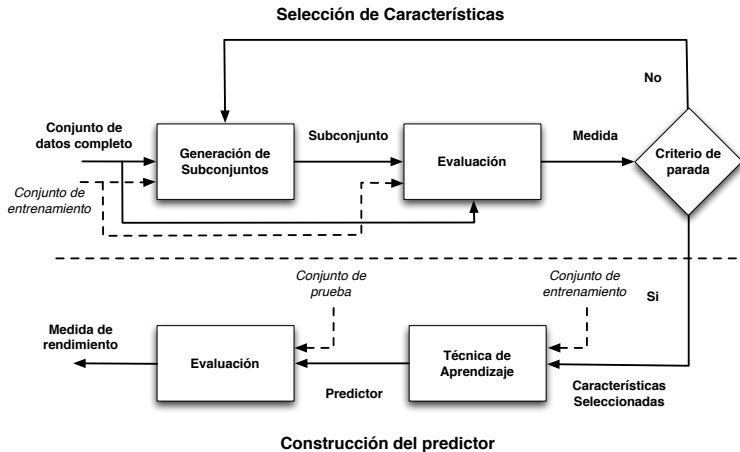
Subconjunto óptimo

- El **subconjunto óptimo** de características es el subconjunto mínimo que permite construir una hipótesis consistente con los datos de entrenamiento.
 - Podemos utilizar todos los datos o sólo los de entrenamiento.
 - El subconjunto se puede sobreajustar a los datos de entrenamiento.
- Sea F el conjunto de todas las características y C el conjunto de clases.
- El subconjunto óptimo es el conjunto mínimo de características G tal que $P(C|G)$ es igual, o lo más próxima posible, a $P(C|F)$.
 - La selección del conjunto óptimo se puede realizar sobre todo el conjunto de datos o sólo sobre los de entrenamiento.

Selección de características: Fases

- El proceso de selección de características se desarrolla en dos fases:
 - 1 Generación de subconjuntos de características. Se van generando los distintos subconjunto de características candidatos a ser el subconjunto óptimo.
 - 2 Se evalúan cada uno de los subconjuntos candidatos hasta que se cumpla algún criterio de parada.
- Finalmente, se procede a generar un modelo teniendo en cuenta sólo las características seleccionadas.

Selección de características: Fases



Selección de características: Fases

- Dependiendo de si se lleva a cabo una generación de subconjuntos y posterior evaluación de los mismos:
 - 1 **Métodos basados en selección de conjuntos de características**, que son aquellos que incluyen alguna estrategia para la generación y evaluación de características.
 - 2 **Métodos de evaluación de características (“Feature ranking” o “rankers”)** que son aquellos que no realizan ninguna generación de subconjuntos, evaluando sólo las variables, estableciendo un “ranking” de acuerdo con alguna medida.

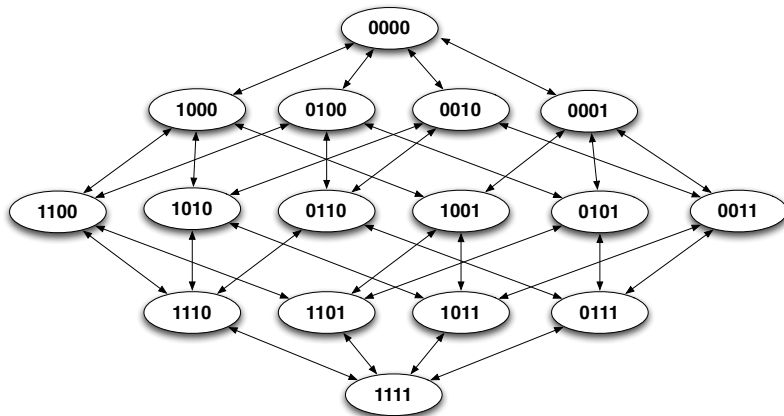
Generación de subconjuntos

- El proceso de selección de características se puede plantear como un proceso de búsqueda en el espacio de posibles subconjuntos de características.
- Una búsqueda exhaustiva, en la mayoría de los casos, no es computacionalmente posible.
 - con $n = 12$ características tendríamos 4096 subconjuntos y para $n = 100$ características tendríamos 10^{100} .
 - Incluso en el caso de que estuviéramos buscando exactamente un subconjunto de tamaño m el número de posibles subconjuntos sería:

$$\binom{n}{m} = \frac{n!}{(n-m)!m!}$$

que puede ser mucho menor que 2^n , pero desde el punto de vista computacional muy grande.

Generación de subconjuntos



Ejemplo de espacio de estado para 4 características.

Generación de subconjuntos: Estrategias de búsqueda I

- Se pueden utilizar diferentes estrategias de búsqueda en el espacio de posibles subconjuntos:
 - **Exhaustiva.** Se barre todo el espacio de posibles subconjuntos:
 - Se puede recorrer el espacio en profundidad o en anchura.
 - Sólo es posible para pocas características.
 - Es la única forma de garantizar encontrar el subconjunto óptimo.
 - **Heurística.** Disponen de alguna información sobre qué subconjunto es el más prometedor.
 - No garantizan que el subconjunto encontrado sea el óptimo.
 - Normalmente encuentran una buena solución en un tiempo razonable.

Generación de subconjuntos: Estrategias de búsqueda II

- **Aleatoria.** Se parte de una configuración inicial formada por un conjunto finito de posibles subconjuntos.
 - Mediante una pequeña transformación, se va modificando la configuración inicial para dirigir la búsqueda hasta la solución final.
 - No garantiza la solución óptima.

Generación de subconjuntos: Dirección de Búsqueda

- Para definir una estrategia de búsqueda primero debemos definir **la dirección de búsqueda**:
 - **Hacia delante (Forward)**. Se empieza con el conjunto vacío y se van añadiendo una característica no seleccionada cada vez.
 - **Hacia atrás (Backward)**. Se empieza con todas las características y se va eliminando una característica cada vez.
 - **Bidireccional**. Se comienza por los dos extremos del espacio de búsqueda y se realiza de forma paralela una búsqueda hacia delante y otra hacia detrás.

Evaluación

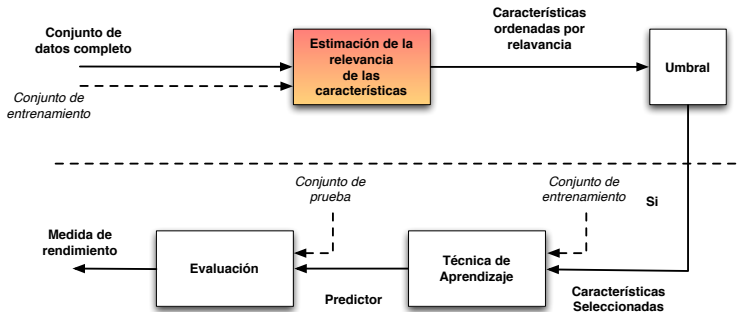
- Dependiendo de cómo se realice la evaluación de cada uno de los subconjuntos tenemos:
 - **Métodos basados en filtros**, que evalúan la relevancia de las características sin tener en cuenta la tarea que se va a realizar con los datos.
 - **Métodos basados en envoltura (wrappers)**, en los que cada subconjunto candidato es evaluado a través de la eficacia de un predictor. Por lo tanto, se tiene en cuenta la tarea que se va a realizar.
 - **Métodos empotrados (embedded)**, en los que el proceso de selección de variables está integrado en la técnica de construcción del predictor.

Evaluadores de características

- La característica básica de este tipo de métodos es que eliminan la fase de búsqueda de subconjuntos.
- Por lo tanto, lo único que hacen es evaluar cada variable de acuerdo con una determinada medida y establecer un "ranking" de las mismas.
- Una vez obtenida la lista, se seleccionan aquellas características que están mejor clasificadas.
- Este tipo de técnicas son muy rápidas y fácilmente escalables.
- Tiene el inconveniente de que en algunos casos es difícil establecer el umbral de corte.

Evaluadores de características

Selección de Características



Construcción del predictor

Esquema general de un método de un filtro de tipo ranker

Evaluadores de características: Medidas

- Las medidas utilizadas para evaluar características se pueden clasificar en:
 - **Medidas de evaluación por pares**, que son aquellas técnicas que evalúan la dependencia de las variables con la variable objetivo.
 - Este tipo de medidas no nos permite evaluar la interacción de grupos de variables respecto a la variable objetivo → **análisis univariable**
 - **Medidas de evaluación simultáneas**, que son aquellas que realizan la evaluación o "ranking" de forma simultánea para todas las variables.
 - Este tipo de medidas si permiten la evaluación de la interacción entre variables → **análisis multivariable**

Evaluadores de características: Medidas de evaluación por pares

- Las medidas de evaluación por pares podemos encontrar medias se pueden agrupar en:
 - **Basadas en correlación**, que miden la correlación entre cada variable y la variable objetivo.
 - **Basadas en incertidumbre**, que se apoyan en medidas típicas de la teoría de la información como la entropía. La más característica es la ganancia de información.
 - **Basadas en test de hipótesis** mediante los cuales se obtienen los "p-values" que pueden ser usados para clasificar las variables.
 - **Basadas en el poder discriminativo** utilizando un modelo de una sola variable y estimando su error

Evaluadores de características: Medidas de evaluación simultáneas

- Las medidas de evaluación simultáneas permiten analizar la interacción y dependencia entre variables.
- El ejemplo más característico de este tipo de técnicas es la familia Relief [[Kira and Rendell, 1992](#)].
 - La relevancia de las variables se calcula basándose en medidas de distancia.
 - Devuelve una medida entre $[-1, 1]$ donde 1 indicaría la mayor relevancia de la variable.

Evaluadores de características: Relief

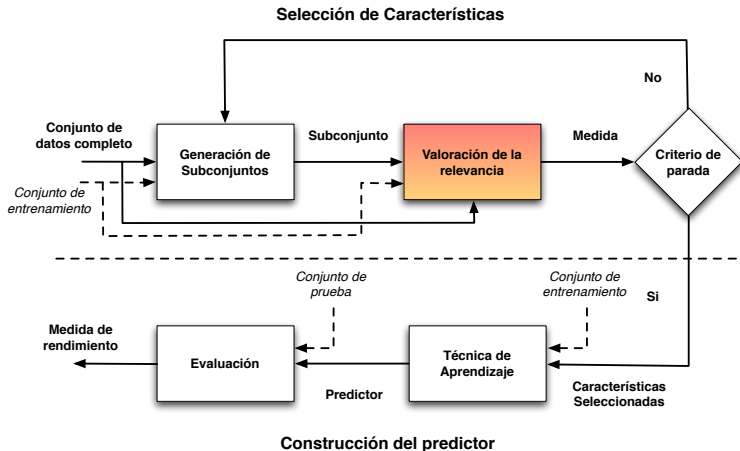
Algoritmo Relief

- 1: **Entrada:** Dataset \mathbf{D} , Número de iteraciones T
 - 2: **Salida:** Pesos de las características, \mathbf{W}
 - 3: $\mathbf{W} \leftarrow 0$
 - 4: **para** $t = 1$ to T **hacer**
 - 5: **para** cada instancia $x_i \in \mathbf{D}$ **hacer**
 - 6: Encontrar el "nearest hit" , nh y el "nearest miss" nm de x_i
 - 7: **para** cada variable j **hacer**
 - 8: Actualizar su peso: $w_j = w_j - \frac{1}{T} \cdot (|x_{ij} - nh_j|^2 - |x_{ij} - nm_j|^2)$
 - 9: **fin para**
 - 10: **fin para**
 - 11: **fin para**
 - 12: Normalizar los pesos \mathbf{W}
 - 13: **devolver** \mathbf{W}
-

Filtros

- Los métodos basados en filtros evalúan la relevancia de las características teniendo en cuenta sólo las propiedades intrínsecas de las mismas, ya que no tienen en cuenta la tarea que se va a realizar.
- Son independientes de la técnica de clasificación/regresión que a realizar.
 - La búsqueda en el espacio de subconjunto de características está desacoplado del de las hipótesis.
- No se ven sesgados por la influencia de una determinada técnica de clasificación/regresión.

Filtros



Esquema general de un método de selección basado en filtros

Filtros: Medidas de relevancia

- Ejemplos de medidas de relevancia pueden ser
 - **Ratio de inconsistencias.** Una inconsistencia aparece cuando dos instancias iguales, de acuerdo con el subconjunto de características evaluados, pertenecen a clases distintas.
 - **Medidas basadas en distancias,** que se centran en encontrar el subconjunto de características que mejor separe las diferentes clases. Ejemplos de medidas ya han sido analizadas en el capítulo de técnicas de clustering.

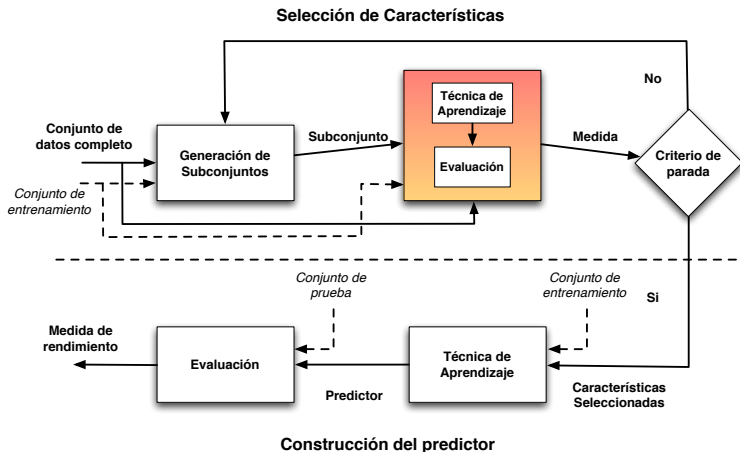
Filtros: Algoritmos representativos

- FOCUS [Almuallim and Dietterich, 1994]: búsqueda exhaustiva con medida de consistencia.
- Cluster-based Feature Selection Approach [Covões *et al.*, 2009]: k-medias basado en medidas de correlación en el espacio de características, utilizando el criterio del índice silueta como criterio para elegir el número óptimo de clusters.
- CFS (Correlation-Based Feature Selection) [Hall, 1999]: búsqueda heurística (ramificación y poda) con una medida basada en la correlación.

"Wrappers"

- Este tipo de métodos, la evaluación de cada uno de los subconjuntos candidatos se realiza mediante la construcción de un clasificador (o regresor).
 - Por lo tanto, como medida de evaluación utilizan la capacidad predictiva del clasificador.
- De esta forma, se selecciona el subconjunto de características que produce el mejor clasificador.
- Ventajas:
 - Producen mejores resultados al estar orientados al problema de clasificación.
- Desventajas:
 - Tienen un mayor riesgo de sobreajuste que los filtros.
 - Son más costosos computacionalmente.

"Wrappers"



Esquema general de un método de un wrapper

"Wrappers": esquema con eliminación recursiva

Algoritmo Eliminación recursiva de características

- 1: Construir un modelo con todas las características
 - 2: Evaluar el modelo
 - 3: Calcular la relevancia de las características
 - 4: Crear una lista con las características ordenadas de mayor a menor relevancia.
 - 5: **para** $size = n$ to 1 **hacer**
 - 6: Crear un conjunto S_{size} con las $size$ características más relevantes
 - 7: Construir un modelo utilizando las características S_{size}
 - 8: Evaluar el modelo
 - 9: [Opcional] Recalcular la relevancia de las características
 - 10: **fin para**
 - 11: Crear una lista con todos los S_i y el resultado de la evaluación
 - 12: Determinar el subconjunto óptimo S_{opt}
-

"Wrappers": esquema con eliminación recursiva

- El mismo esquema se puede adaptar para una búsqueda hacia delante de subconjuntos.
 - Se empieza por un conjunto de tamaño 1.
 - Se van agregando las características más relevantes.
- La principal desventaja de este tipo de métodos, a parte de su coste computacional, es el sobreajuste.
 - Para evitarlo, hay esquemas que incluyen un bucle externo para llevar a cabo un remuestreo.

"Wrappers": eliminación recursiva con remuestreo

Algoritmo Eliminación recursiva de características con remuestreo

- 1: **para** *Cada iteración de remuestreo* **hacer**
 - 2: Crear los conjunto de entrenamiento E y prueba T
 - 3: Construir un modelo sobre E con todas las características
 - 4: Evaluar el modelo en T
 - 5: Calcular la relevancia de las características
 - 6: Crear una lista con las características ordenadas de mayor a menor relevancia.
 - 7: **para** $size = n$ to 1 **hacer**
 - 8: Crear un conjunto S_{size} con las $size$ características más relevantes
 - 9: Construir un modelo utilizando las características S_{size}
 - 10: Evaluar el modelo
 - 11: [Opcional] Recalcular la relevancia de las características
 - 12: **fin para**
 - 13: **fin para**
 - 14: Crear una lista con todos los S_i y el resultado de la evaluación
 - 15: Determinar el subconjunto óptimo S_{opt}
 - 16: Crear un modelo con las variables S_{opt} y con el conjunto de entrenamiento original.
-

"Wrappers": Algoritmos representativos

- OBLIVION [Langley and Sage, 1994]: búsqueda voraz y árboles de decisión.
- RFE+SVM [Guyon *et al.*, 2002]: eliminación recursiva de características y SVM.
- FFE + NNets [Goutte, 1997]: búsqueda hacia delante y redes neuronales.
- GA + C4.5 [Abbasimehr and Alizadeh, 2013]: búsqueda aleatoria y C4.5.
 - También se pueden encontrar muchos ejemplos basados en colonias de hormigas y enfriamiento simulado.
- También es posible implementar modelos de tipo "ranker" con predictores que calculen la relevancia de las características.

Métodos embebidos

- En los métodos embebidos, la selección de características está integrada en el proceso de construcción del modelo.
 - Al mismo tiempo que se ajustan los parámetros internos, la técnica de aprendizaje selecciona las características más relevantes.
- Computacionalmente, al igual que los métodos basados en filtros, son menos costosos que los métodos “wrappers”, con la ventaja sobre los filtros de que, además, generan el modelo.
- Al igual que ocurre con los métodos “wrappers”, tienen la ventaja de que la selección de características está sesgada hacia una mejor eficiencia del proceso de clasificación/regresión.

Métodos embebidos

Algunos ejemplos:

- Técnicas basadas en árboles de decisión, como por ejemplo, C5.0, Random Forest, Gradient Boosting.
 - Este tipo de técnicas permite analizar la interacción entre características (métodos multivariantes).
- Modelos con regularización, como por ejemplo LASSO o Elastic Nets
 - Estas técnicas de regularización se utilizan normalmente en modelos lineales (SVM, regresión lineal o logística) penalizando los coeficientes de las características menos significativas.
- La mayoría de este tipo de técnicas permiten obtener un ranking de características.
 - Importancia de las variables en las técnicas basadas en árbol.
 - Los valores de los coeficientes en las técnicas basadas en regularización.

Selección de características: conclusiones I

- En la actualidad, es crucial la aplicación de técnicas de reducción de la dimensionalidad antes de abordar la creación de modelos predictivos.
- Estas técnicas se pueden agrupar en dos grandes grupos: técnicas de **extracción de características** y **selección de características**.
- En este capítulo nos hemos centrado en las técnicas de selección características, cuyo objetivo es encontrar un conjunto óptimo de características de acuerdo con algún criterio.

Selección de características: conclusiones II

- Dependiendo del criterio elegido las técnicas pueden ser de tipo filtro o de tipo envoltura (wrapper):
 - Las técnicas de tipo filtro sólo se apoyan en las propiedades intrínsecas de las características.
 - Las técnicas de tipo wrappers se basan en la construcción de clasificadores (o regresores) para evaluar la idoneidad de los conjuntos de características.
- Un aspecto importante en ambos tipos de técnicas es la técnica de búsqueda utilizada para recorrer el espacio de búsqueda de posibles subconjuntos de características.
 - Búsquedas exhaustivas, heurísticas o aleatorias.

Referencias I



H. Abbasimehr and S. Alizadeh.

A novel genetic algorithm based method for building accurate and comprehensible churn prediction models.

International Journal of Research in Industrial Engineering, 2(4):1–14, 2013.



Hussein Almuallim and Thomas G Dietterich.

Learning boolean concepts in the presence of many irrelevant features.

Artificial Intelligence, 69(1):279–305, 1994.



Ethem Alpaydin.

Introduction to machine learning.

MIT press, 2014.



Krzysztof J Cios, Witold Pedrycz, and Roman W Swiniarski.

Data mining methods for knowledge discovery, volume 458.

Springer Science & Business Media, 2012.

Referencias II



Thiago F. Covões, Eduardo R. Hruschka, Leandro N. de Castro, and Átila M. Santos.

A cluster-based feature selection approach.

In Emilio Corchado, Xindong Wu, Erkki Oja, Álvaro Herrero, and Bruno Baruque, editors, *Hybrid Artificial Intelligence Systems*, pages 169–176, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.



Richard O Duda, Peter E Hart, and David G Stork.

Pattern classification.

John Wiley & Sons, 2012.



C. Goutte.

Extracting the relevant decays in time series modelling.

In *Proceedings of the VII IEEE Workshop, Neural Networks for Signal Processing*, 1997.



Isabelle Guyon and André Elisseeff.

An introduction to feature extraction.

In Isabelle Guyon, Masoud Nikraves, Steve Gunn, and LotfiA. Zadeh, editors, *Feature Extraction*, volume 207 of *Studies in Fuzziness and Soft Computing*, pages 1–25. Springer Berlin Heidelberg, 2006.

Referencias III



Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik.
Gene selection for cancer classification using support vector machines.
Machine Learning, 46(1):389–422, 2002.



Mark A Hall.
Correlation-based feature selection for machine learning.
PhD thesis, The University of Waikato, 1999.



Kenji Kira and Larry A Rendell.
A practical approach to feature selection.
In *Proceedings of the ninth international workshop on Machine learning*, pages
249–256, 1992.



Vipin Kumar and Sonajharia Minz.
Feature selection: A literature review.
Smart Computing Review, 4(3):211–229, June 2014.



Langley and Stephanie Sage.
Oblivious decision trees and abstract cases.
In *Proc. AAAI-94 Workshop on case-based reasoning*, pages 113–117, 1994.

Referencias IV



Yvan Saeys, Iñaki Inza, and Pedro Larrañaga.

A review of feature selection techniques in bioinformatics.

bioinformatics, 23(19):2507–2517, 2007.