

EVIDENCIAS EN PEDIATRÍA

Toma de decisiones clínicas basadas en las mejores pruebas científicas

www.evidenciasenpediatria.es

Fundamentos de medicina basada en la evidencia

Principales algoritmos de aprendizaje automático. Parte 1

Molina Arias M

Servicio de Gastroenterología Pediátrica. Hospital Universitario Infantil La Paz. Madrid. España.

Correspondencia: Manuel Molina Arias, mma1961@gmail.com

Palabras clave en español: aprendizaje automático; máquinas de vectores de soporte; regresión lineal; regresión logística.

Palabras clave en inglés: machine learning; supported vector machines; logistic regression; linear regression.

Fecha de recepción: 20 de enero de 2026 • **Fecha de aceptación:** 30 de enero de 2026

Fecha de publicación del artículo: 11 de febrero de 2026

Evid Pediatr. 2025;22:9.

CÓMO CITAR ESTE ARTÍCULO

Molina Arias M. Principales algoritmos de aprendizaje automático. Parte 1. Evid Pediatr. 2025;22:9.

Para recibir Evidencias en Pediatría en su correo electrónico debe darse de alta en nuestro boletín de novedades en
<http://www.evidenciasenpediatria.es>

Este artículo está disponible en: <http://www.evidenciasenpediatria.es/EnlaceArticulo?ref=2025;22:9>.

©2005-26 • ISSN: 1885-7388

Este es un artículo Open Access bajo la licencia

CC BY-NC-ND (Reconocimiento-No comercial-Sin obras derivadas): <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>

Principales algoritmos de aprendizaje automático.

Parte 1

Molina Arias M

Servicio de Gastroenterología Pediátrica. Hospital Universitario Infantil La Paz. Madrid. España.

Correspondencia: Manuel Molina Arias, mma1961@gmail.com

INTRODUCCIÓN

En artículos previos de esta serie de *Fundamentos de Medicina Basada en la Evidencia* describimos los dos tipos principales de aprendizaje automático, supervisado y no supervisado, así como los parámetros y el funcionamiento general de estos algoritmos. En este artículo y los siguientes hablaremos de los algoritmos más utilizados, describiendo de forma somera su estructura, su funcionamiento general y sus principales ventajas e inconvenientes.

Comenzaremos con los algoritmos de aprendizaje supervisado, que pueden utilizarse para la predicción tanto de variables cuantitativas (algoritmos de regresión) como cualitativas (algoritmos de clasificación).

REGRESIÓN LINEAL

Dentro del arsenal de métodos de estadística y de aprendizaje automático, la regresión lineal es, probablemente, la más utilizada para comprender la relación cuantitativa entre variables. Este es un algoritmo de aprendizaje supervisado de regresión.

En su forma más elemental, la regresión lineal simple, permite modelar mediante una línea recta cómo una única variable independiente (el predictor) influye sobre una variable dependiente continua (el resultado). Podemos representar el modelo con la ecuación siguiente:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

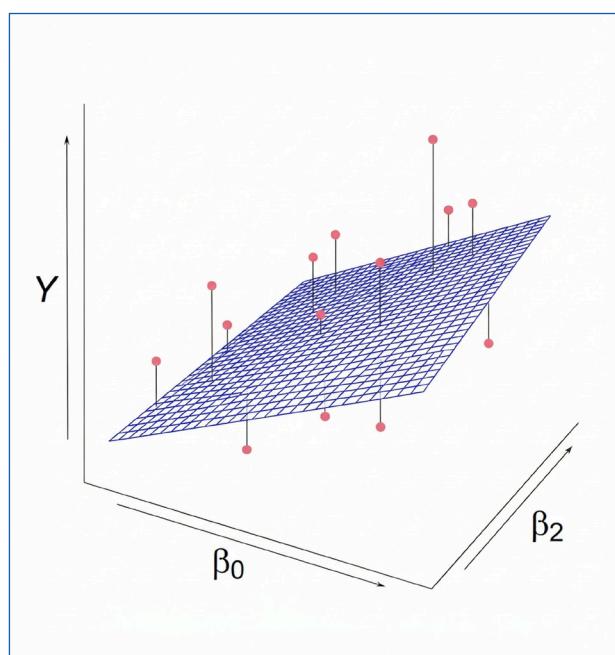
Imaginemos un escenario clínico habitual: un investigador desea cuantificar cómo el incremento en la dosis de un nuevo antiinflamatorio (variable X) se traduce directamente en la reducción de un biomarcador inflamatorio específico, como la proteína C reactiva (variable Y). Aquí, el modelo traza una línea recta que intenta predecir el valor del biomarcador basándose exclusivamente en la dosis administrada.

Sin embargo, la fisiología rara vez es unidimensional. Para capturar la complejidad biológica de nuestros escenarios, debemos extender este concepto hacia la regresión lineal múltiple.

Esta evolución del modelo permite la inclusión simultánea de múltiples predictores para ajustar el resultado, lo que posibilita aislar el efecto real del tratamiento controlando por factores de confusión, ofreciendo una visión más ajustada de la realidad. En este caso no se modela una recta, sino un plano o un hiperplano multidimensional, según el número de variables independientes o predictores que incluyamos en el modelo (**Figura 1**).

Conceptualmente, la regresión lineal busca ajustar un modelo lineal a los datos. Para encontrar la “mejor” línea o plano que se ajusta a los datos, utiliza el método de mínimos cuadrados, que minimiza la suma de los errores al cuadrado entre los valores reales y los predichos por el modelo. Alternativamente, los coeficientes del modelo pueden calcularse también mediante el entrenamiento utilizando métodos de optimización, como el gradiente descendente, especialmente cuando el número de variables o el tamaño del conjunto de datos es elevado. Las métricas de error más comunes para evaluar este ajuste son el error cuadrático medio (MSE) y la raíz del error cuadrático medio (RMSE). Las métricas de desempeño

Figura 1. Representación de un plano de regresión lineal múltiple (dos variables predictoras)



son bien conocidas, siendo las más empleadas el coeficiente de determinación (R^2) y, con menos frecuencia, el error estándar de los residuos.

A diferencia de los modelos de “caja negra” del aprendizaje profundo, los modelos lineales son paramétricos y no requieren el ajuste de hiperparámetros complejos; no obstante, la selección rigurosa de las variables clínicas que se incluyen en el modelo es una decisión crítica que determina su validez.

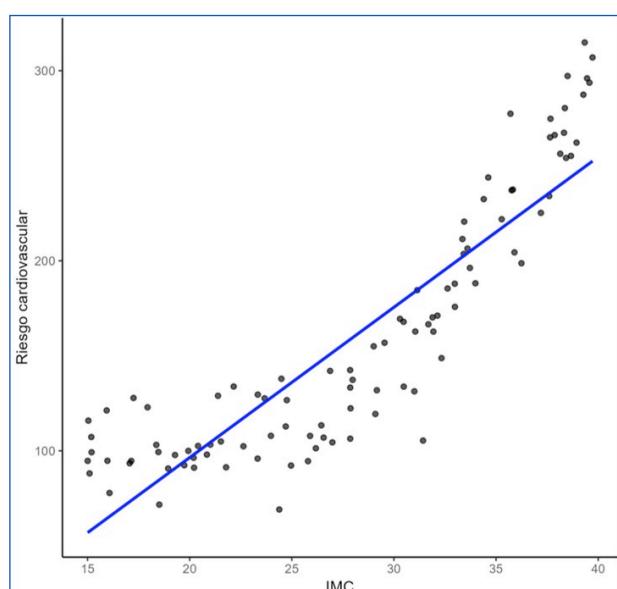
Interpretabilidad frente a linealidad

La gran fortaleza de la regresión lineal reside en su alta interpretabilidad clínica. A diferencia de algoritmos más opacos, permite cuantificar exactamente cuánto influye cada factor (por ejemplo, “por cada kg de peso adicional, el biomarcador disminuye en X unidades”), lo cual es fundamental para validar el juicio clínico y fomentar la confianza en las decisiones terapéuticas. Además, una vez entrenado, el modelo ofrece una capacidad de predicción casi instantánea.

No obstante, debemos ser cautelosos con su utilización: la principal limitación es que este método asume que la relación entre los predictores y el resultado es, en efecto, lineal. Si la respuesta biológica es curva o exponencial, la regresión lineal simplificará excesivamente la realidad, introduciendo sesgo (**Figura 2**). Asimismo, aunque generalmente ágil, el proceso de entrenamiento puede volverse computacionalmente costoso si se trabaja con conjuntos de datos masivos (*Big Data* genómico o poblacional), lo que requiere una consideración cuidadosa de los recursos disponibles.

Como extensión de los métodos de regresión lineal, podemos encontrar otras técnicas más complejas, como la

Figura 2. Ajuste lineal a datos no lineales. El modelo realizará predicciones inexactas



regresión polinómica (introduce términos no lineales al modelo mediante transformaciones polinómicas de una variable), la regresión con interacciones (incluye términos de interacción para modelar el efecto combinado de dos o más variables), la regresión por componentes principales (usa análisis de componentes principales para reducir la dimensionalidad antes de la regresión) o la regresión por splines (ajusta funciones no lineales a los datos dividiendo el rango de la variable independiente en intervalos y ajustando polinomios de bajo grado en cada intervalo), entre otras.

Métodos de regularización de la regresión

En el contexto del uso de algoritmos de aprendizaje automático, es frecuente encontrarse con bases de datos donde el número de variables (datos genómicos, biomarcadores...) es muy elevado e, incluso, supera al número de registros del conjunto de datos. En estos casos, los modelos de regresión tradicionales corren el riesgo de caer en el sobreajuste (*overfitting*): el modelo “memoriza” el ruido y las peculiaridades de la muestra de entrenamiento en lugar de aprender las relaciones biológicas reales, perdiendo así su capacidad de generalizar cuando se le enfrenta a datos nuevos.

Para mitigar este efecto, se recurre con frecuencia a las técnicas de regularización, que introducen un “coste” o penalización a la complejidad del modelo, haciendo que los valores de los coeficientes de la recta o plano de ajuste sean lo más bajos posible. Esto fuerza al algoritmo a priorizar solo aquellas relaciones que son verdaderamente fuertes y consistentes, filtrando el ruido estadístico.

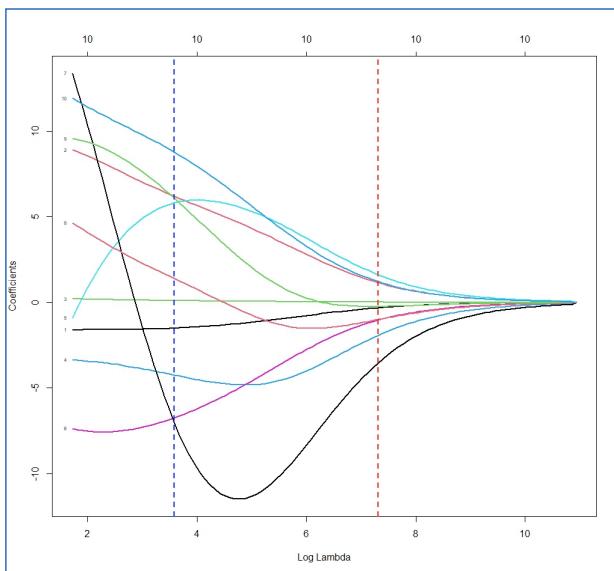
Existen dos tipos fundamentales de regularización

La regresión Ridge (L2) añade una penalización al método de los mínimos cuadrados basada en el cuadrado de los coeficientes. Su efecto es el de “encoger” (*shrinkage*) todos los coeficientes hacia cero de manera proporcional, pero sin anularlos por completo. Es la elección metodológica ideal cuando asumimos que todas las variables contribuyen al resultado, aunque sea mínimamente, o cuando existe alta colinealidad (variables muy correlacionadas entre sí, como peso y talla).

Por otra parte, la regresión Lasso (L1), a diferencia de la Ridge, penaliza basándose en el valor absoluto de los coeficientes, con lo que puede forzar a que algunos coeficientes sean exactamente cero (**Figura 3**). En la práctica, esto actúa como un mecanismo de selección automática de variables, descartando las superfluas y produciendo modelos más simples y fáciles de interpretar.

El control de la intensidad de la penalización se realiza mediante el hiperparámetro lambda (λ), cuyo valor óptimo debe encontrarse durante el entrenamiento, habitualmente usando

Figura 3. Representación gráfica de una regularización L1 o Lasso. En el eje de abscisas se muestran los valores de lambda (en escala logarítmica) y en el de ordenadas los valores de los coeficientes del modelo de regresión. Cada coeficiente, correspondiente a una variable predictora diferente, se muestra en un color diferente. Puede observarse como al aumentar el valor de lambda (la penalización), los valores de los coeficientes disminuyen hasta alcanzar el valor cero



técnicas de validación cruzada. Valores más elevados imponen una mayor penalización, mientras que valores próximos a cero hacen que el modelo se asemeje más a la regresión lineal estándar.

Estas dos técnicas pueden utilizarse de forma conjunta en lo que se denomina regularización en red elástica, en la que la intensidad de cada modalidad se regula mediante otro hiperparámetro denominado alfa (α).

Por último, es fundamental comprender que este principio de “penalizar la complejidad” trasciende la regresión lineal. Es un pilar transversal en el aprendizaje automático que reaparece en otros algoritmos, como la regresión logística o los árboles de decisión con métodos de ensemble.

REGRESIÓN LOGÍSTICA

A pesar de conservar el nombre de “regresión” (una herencia histórica), la regresión logística es un algoritmo de aprendizaje supervisado de clasificación. Su función es estimar la probabilidad de que una observación pertenezca a una categoría específica. Cuando esta variable objetivo es dicotómica hablamos de regresión logística binaria, que puede ser simple o múltiple, en función del número de variables independientes o predictoras. Cuando la variable objetivo tiene más de dos categorías, se utiliza la regresión logística multinomial.

La curva sigmoide y la máxima verosimilitud

A diferencia de la línea recta de la regresión lineal (que teóricamente podría extenderse hacia el infinito), la regresión logística utiliza la función sigmoide (o logística). Esta función transforma cualquier valor de entrada en un número estrictamente comprendido entre 0 y 1, dibujando una curva característica en forma de “S”.

Debido a esta no linealidad, ya no es correcto usar el método de mínimos cuadrados para entrenar el modelo. En su lugar, se recurre a la estimación de máxima verosimilitud. En términos sencillos, el algoritmo busca iterativamente los coeficientes que hacen que la probabilidad predicha para los eventos observados sea lo más alta posible. La función de coste que guía este proceso de optimización se denomina *Log-loss* (pérdida logarítmica o entropía cruzada), la cual penaliza fuertemente al modelo si predice una probabilidad baja para un evento que sí ocurrió (o viceversa).

Métricas de desempeño

Las métricas de desempeño son conocidas en nuestro entorno por su utilización habitual para la valoración de las pruebas diagnósticas, aunque existe alguna más que es más específica del mundo de la ciencia de datos y el aprendizaje automático.

Las más conocidas, en las que no vamos a entrar en detalle, son sensibilidad y especificidad, valores predictivos positivo y negativo, y curva ROC con la determinación del área bajo la curva. Además, podemos encontrarnos otras dos métricas que se utilizan con frecuencia con estos algoritmos: la exactitud y la puntuación F1 (F1-Score).

La exactitud es la proporción total de predicciones correctas. Esta métrica debe utilizarse con precaución en el ámbito de la pediatría, donde muchas patologías son poco frecuentes (clases desbalanceadas). Si una enfermedad tiene una prevalencia muy baja, un modelo que siempre prediga “sano” tendrá una exactitud muy elevada, pero una utilidad clínica nula (tendrá poca o nula capacidad para detectar a los enfermos).

El F1-Score es la media armónica entre la sensibilidad y el valor predictivo positivo (¡ojo! en inglés, suelen denominarse *recall* y *precision*, respectivamente, en ciencia de datos). Funciona como un resumen del equilibrio del modelo, siendo especialmente útil cuando buscamos un balance entre no perder casos y no sobrediagnosticar. Combinada con el área bajo la curva ROC, puede ser muy útil en los casos de clasificación con categorías muy desbalanceadas.

Estas métricas de desempeño son comunes al resto de algoritmos de clasificación que veremos en próximas secciones.

Por último, al igual que ocurría con la regresión lineal, existen extensiones de las técnicas de regresión logística binaria y multinomial, como la regresión logística regularizada (vistas al hablar de regresión lineal), la regresión logística con splines o no lineal (introduce splines o funciones base para modelar relaciones no lineales entre las variables), la regresión logística penalizada bayesiana (asigna distribuciones previas sobre los coeficientes) y la regresión logística para datos desbalanceados (reajusta los pesos de las clases y aplica técnicas de sobremuestreo o submuestreo, como SMOTE), entre otras.

MÁQUINAS DE VECTORES DE SOPORTE

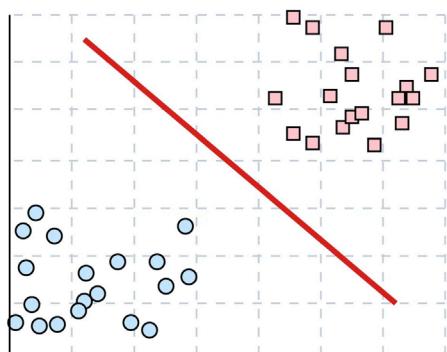
Los dos algoritmos que hemos visto hasta ahora intentan minimizar el error total a toda costa. Sin embargo, en la práctica clínica, a veces no necesitamos una precisión milimétrica en cada punto, sino un modelo que capture la tendencia general sin distraerse por variaciones menores. En estos casos puede ser útil recurrir a las máquinas de vectores de soporte (SVM), más conocidas por su aplicación como algoritmos de clasificación, pero que pueden utilizarse también para regresión.

Clasificación con máquinas de vectores de soporte

Mientras que la regresión logística calcula probabilidades, las SVM son especialistas en geometría. Este algoritmo es un potente clasificador cuyo objetivo es encontrar el “hiperplano” (una línea en 2D, un plano en 3D o una superficie en dimensiones superiores) que separe de forma óptima dos clases de pacientes ([Figura 4](#)).

Figura 4. Separación lineal mediante hiperplanos. Un hiperplano tiene siempre una dimensión menos que la del espacio en el que se representa. En un espacio bidimensional, un hiperplano es una recta, en uno tridimensional, es un plano de dos dimensiones. El objetivo es que todos los puntos se encuentren a uno u otro lado del hiperplano. En la práctica, esto suele realizarse en espacio multidimensionales, imposibles de representar de forma gráfica

En un espacio bidimensional, un hiperplano es una recta



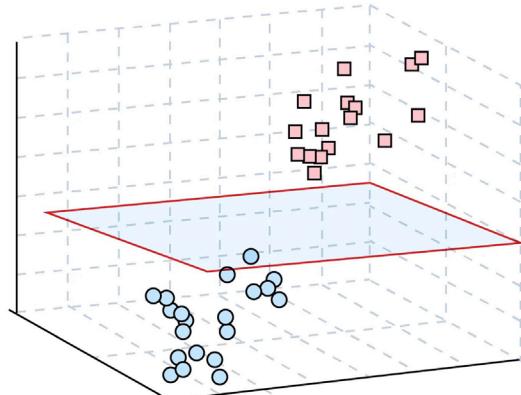
Lo que distingue a la SVM de otros clasificadores es su búsqueda del margen máximo. Las SVM no se conforman con cualquier línea que separe los grupos, sino que buscan la línea que pase lo más lejos posible de los elementos de ambos grupos o categorías. Aquí surge el concepto que da nombre al algoritmo: los vectores de soporte. Estos son los casos “límite”, los elementos cuya clasificación es más difícil porque sus características son ambiguas y están justo al borde de la frontera de decisión. Curiosamente, el algoritmo no depende de los casos obvios (los puntos lejanos); todo el modelo se construye y se define exclusivamente basándose en estos casos difíciles y más próximos a la frontera ([Figura 5](#)).

El “truco del kernel”: separar lo inseparable

Existen ocasiones en que los grupos no son separables mediante una línea recta o un hiperplano sin cometer un error excesivamente alto. En estos casos, en lugar de forzar una separación en el espacio original, el algoritmo utiliza una función matemática (kernel) para proyectar los datos a una dimensión superior (3D o más). Es lo que se conoce como el *kernel trick*.

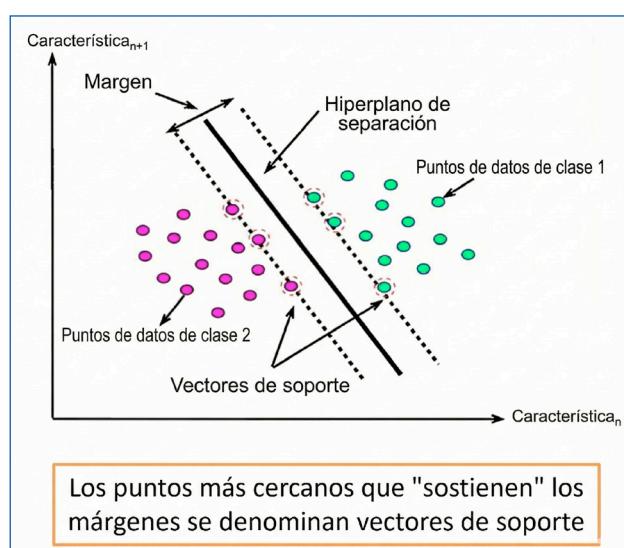
Podemos ver una analogía visual en la [Figura 6](#). Si tenemos puntos rojos y verdes mezclados en una hoja de papel (2D) y no podemos separarlos, el kernel “levanta” los puntos rojos hacia arriba. Ahora, podemos deslizar una hoja de papel (un hiperplano) entre ellos para separarlos perfectamente. Al volver a mirar el papel original desde arriba, esa separación recta se ve como una curva compleja y no lineal.

En un espacio tridimensional, un hiperplano es un plano



Todo punto del hiperplano cumple la siguiente condición: $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p = 0$

Figura 5. Representación gráfica de una máquina de vectores de soporte para clasificación en dos dimensiones (dos variables o características predictoras). Se busca un hiperplano (en este caso, una recta) que separe los datos de las dos categorías de la variable objetivo, con un margen de seguridad lo más amplio posible. Los puntos que definen este margen son los denominados vectores de soporte



Ajuste y evaluación del desempeño

Los hiperparámetros más relevantes son el tipo de kernel (lineal, polinomial o radial) y el parámetro de regularización C, que decide cuánto error de clasificación estamos dispuestos a tolerar para mantener el margen amplio.

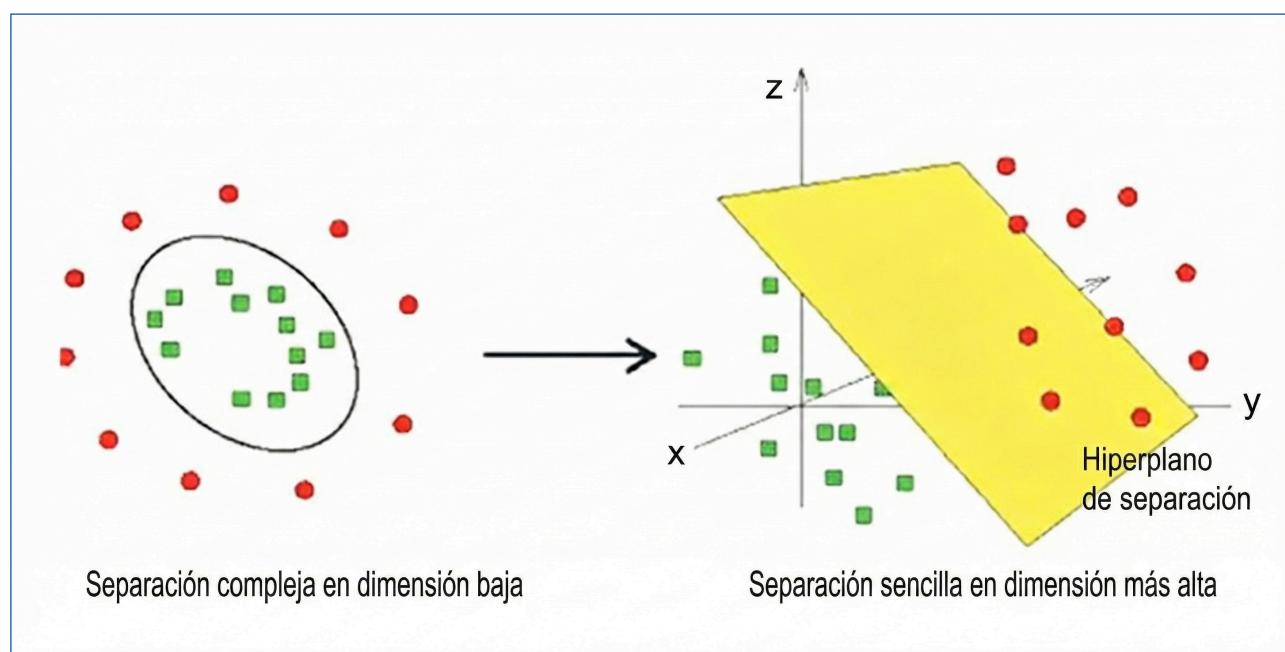
Para validar estos modelos en la clínica, recurriremos a las métricas estándar de clasificación de las que hablamos anteriormente.

Máquinas de vectores de soporte para regresión

Las SVM pueden utilizarse también para la predicción de variables continuas. Al contrario que la regresión lineal clásica, que trata de minimizar cualquier desviación de la predicción, por mínima que sea, las SVM introducen el concepto de un margen de tolerancia de anchura épsilon (ϵ). Los errores de predicción que caen fuera de este rango no se penalizan. Los puntos que caen fuera del margen serían los equivalentes de los vectores de soporte de las SVM de clasificación.

Esta mecánica hace que estos algoritmos sean excepcionalmente robustos frente a pequeños ruidos o valores atípicos (outliers) moderados, ya que no intentan penalizar el error de cada punto individual.

Figura 6. Representación gráfica del efecto de separación de la función del kernel. Los datos no pueden separarse linealmente en dos dimensiones, pero sí es posible hacerlo “elevándolos” a una dimensión mayor. La frontera de separación en el plano ya no es una recta, sino una forma no lineal más compleja



Los hiperparámetros más importantes para controlar el modelo son el kernel, el parámetro ϵ y el parámetro C , que controla la penalización por salirse del margen de tolerancia. Un valor alto de C penaliza mucho los errores durante el entrenamiento (riesgo de sobreajuste), mientras que un C bajo permite más violaciones del margen en favor de una función más suave.

Ventajas e inconvenientes de las SVM

Entre las ventajas están ofrecer buenos resultados incluso con pequeñas cantidades de información, funcionar de forma adecuada con datos no estructurados y en espacios de variables de alta dimensionalidad y su capacidad para resolver problemas de clasificación complejos. Además, existen adaptaciones del algoritmo para realizar clasificación multiclas, no solo binaria.

Sus inconvenientes son la dificultad de elección de la función del kernel más adecuada y su coste computacional para el entrenamiento cuando el conjunto de datos es muy grande.

BIBLIOGRAFÍA

1. Assessment of diagnostic tests. En: Palmas WR, editor. Pocket evidence based medicine: A survival guide for clinicians and students. Nueva York: Springer; 2023. p. 15-33.
2. Ben Braiek H, Khomh F. Machine learning robustness: a primer. Cornell University; 2024. <https://doi.org/10.48550/arXiv.2404.00897>
3. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. 2^a ed. Nueva York: Springer; 2009.
4. James G, Witten D, Hastie T, Tibshirani R, Taylor J. An introduction to statistical learning: with applications in Python. Nueva York: Springer; 2023.
5. Kahouadji N. Comparison of machine learning classification algorithms and application to the Framingham Heart Study. Cornell University; 2024. <https://doi.org/10.48550/arXiv.2402.15005>
6. Khondoker M, Dobson R, Skirrow C, Simmons A, Stahl D. A comparison of machine learning methods for classification using simulation with multiple real data examples from mental health studies. Stat Methods Med Res. 2016;25(5):1804-23.
7. Linear and logistic regression. En: Zumel N, Mount J, editores. Practical Data Science with R. 2^a ed. Shelter Island (NY): Manning Publications Co; 2020. p. 215-73.
8. Linear model selection and regularization. En: James G, Witten D, Hastie T, Tibshirani R, Taylor J, editores. An introduction to statistical learning with applications in Python. Springer-Verlag GmbH; 2023. p. 229-88.
9. Molina Arias M. Inteligencia artificial en Pediatría: de la ciencia ficción a la realidad clínica. Evid Pediatr. 2025;21:11.
10. Molina Arias M. Funcionamiento de los algoritmos de aprendizaje automático. Evid Pediatr. 2025;21:50.
11. Sasaki Y. The truth of the F-measure. Manchester: School of Computer Science, University of Manchester; 2007 [en línea] [consultado el 04/02/2026]. Disponible en www.researchgate.net/publication/268185911_The_truth_of_the_F-measure
12. Swamynathan M. Mastering Machine Learning with Python in Six Steps: A Practical Implementation Guide to Predictive Data Analytics Using Python. Nueva York: Apress; 2017.
13. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc B Methodol. 1996;58(1):267-88.