

## *COLABORACIONES*



# Modelos de valoración automatizada

**Julio Gallego Mora-Esperanza**

*Jefe de Área Regional de Inspección  
Gerencia Regional del Catastro de Madrid*

## La Valoración Tradicional

Una de las mejores definiciones de la valoración inmobiliaria es la que la considera como la “Opinión de un Experto”. No sería, por tanto, un resultado exacto, sino una “opinión” emitida por alguien que no sólo cuenta con conocimientos y experiencia en el sector, sino que además ha construido esta opinión obteniendo y procesando adecuadamente información de mercado y comparando distintas alternativas.

La información disponible del mercado inmobiliario presenta importantes deficiencias derivadas de las características propias de este mercado: producto heterogéneo, problemas de transparencia en el conocimiento de los precios, datos incompletos, información insuficiente sobre determinados productos y zonas, etc.

Es necesaria, por tanto, la presencia de un experto que sea capaz de obtener y procesar esta información deficiente y alcanzar una conclusión razonable.

El trabajo del experto requeriría conocimientos, práctica e intuición. Ha habido incluso sentencias de tribunales que han llegado a indicar que la valoración inmobiliaria no es una ciencia sino que es un arte.

Los métodos de valoración tradicionales son aquellos que se basan fundamentalmente en el criterio de un experto.

Tienen la ventaja de que, al ser los que se han venido realizando hasta la fecha, cuentan con una aceptación alta por parte de particulares, empresas, instituciones, tribunales, etc. Además, alcanzan resultados que se pueden considerar en general bastante buenos en cuanto a la precisión de la valoración.

Sin embargo, también tienen inconvenientes, como son su nivel de subjetividad, entendida no como falta de calidad sino como dificultad para exponer de forma clara y completa el proceso de elaboración, y la baja producción de valores.

## La Valoración Automatizada

En los últimos años se han desarrollado métodos de valoración automatizada, que presentan diferencias significativas respecto a los métodos tradicionales.

En primer lugar, no hay que confundir valoración automatizada con valoración informatizada. Hoy en día se utiliza la informática para todo, y los modelos de valoración tradicionales también están informatizados.

Tampoco hay que confundir valoración automatizada con valoración robotizada. No se trata de que una máquina haga todo el trabajo y obtenga los resultados, ni que una persona sin experiencia en el mercado pueda, con la ayuda de algoritmos matemáticos, realizar el trabajo. El experto sigue siendo necesario.

Los métodos de valoración automatizada se diferencian de los tradicionales en que se basan fundamentalmente en la utilización por parte del experto de técnicas matemáticas para la estimación del valor.

En la actualidad los valoradores suelen disponer de gran cantidad de información de mercado (muy superior a la disponible hace 10 o 15 años). Por otra parte, las técnicas matemáticas de tratamiento de la información para la construcción de modelos (minería de datos, inteligencia artificial etc.) se han desarrollado mucho, y parece conveniente su utilización para procesar la información de mercado disponible y extraer conclusiones que permitan estimar los valores.

El procedimiento sistemático y la utilización de técnicas matemáticas otorgan a la valoración automatizada un carácter más científico y puede considerarse más objetiva y, sobre todo, permite una producción de valores muy superior a la de las valoraciones tradicionales.

Seguidamente se exponen las características fundamentales de estos métodos, así como algunas de las técnicas matemáticas más utilizadas por ellos para la estimación del valor.

## El procedimiento de valoración

Nos vamos a referir en adelante a los procedimientos destinados a la valoración masiva, es decir a las valoraciones completas del parque inmobiliario que realiza la administración, generalmente con finalidad tributaria.

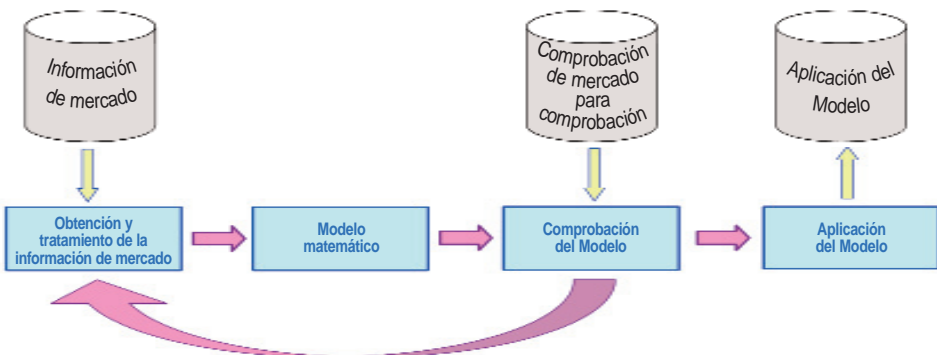
Las etapas para la construcción de un modelo de valoración automatizada son las siguientes.

- Obtención y tratamiento de la información de mercado.
- Elaboración del modelo matemático.
- Comprobación y ajuste del modelo.
- Aplicación del modelo.

En la figura 1 se puede ver un esquema con la secuencia de etapas, y la aportación de información necesaria para realizarlas.

Veamos cada una de estas etapas.

Figura 1. Método de valoración automatizada



## Obtención y tratamiento de la información de mercado

Esta primera etapa es la más importante de todas y la que implica una mayor carga de trabajo. En líneas generales se puede estimar que, de todo el proceso de construcción del modelo, supone un 90% del trabajo.

Las tareas que comprende esta etapa son las siguientes.

- Obtención de la información.
- Calificación de la información.
- Incorporación de variables de situación.
- Tratamiento de la información.
- Selección de variables.
- Normalización de variables.
- Reducción de variables.

Veamos en qué consisten estas tareas.

### Obtención de la información

En primer lugar hay que analizar las fuentes de información disponibles, elegir las más adecuadas y obtener así la información de mercado en origen.

La cuestión es que la información en origen no es procesable directamente. Los modelos matemáticos modelizan lo que se le da, y como la información en origen es bastante deficiente los resultados del modelo dejan mucho que desear. Muchos modelos de valoración han fracasado por no dedicar todo el tiempo necesario a esta etapa. Es verdad que no es la etapa más gratificante, pues no produce ningún resultado, pero la mayor parte del éxito del modelo depende de ella.

Es preciso tratar la información en origen para que el modelo matemático pueda extraer de ella la "verdad oculta" que contiene.

La información de mercado en origen está formada por operaciones u ofertas del mercado inmobiliario, de las que conocemos tanto los precios como todas las características relevantes de los inmuebles, pues el método que se utiliza generalmente para la valoración inmobiliaria es el Método Hedónico, en el que el valor de un bien es función de sus características.

El Método Hedónico implica que el producto debe tener características identificables por compradores y vendedores, que pueden variar, y estas variaciones ocasionan cambios en el precio. En el caso del mercado inmobiliario hablamos de características tales como la superficie, la categoría de la construcción, la antigüedad, etc. Estas características son las variables con las que va a trabajar el modelo matemático.

No existe ningún acuerdo que tenga aceptación generalizada sobre el número de operaciones que se deben tomar para construir un modelo de mercado consistente, pero en general se admite que para la valoración masiva hay que tomar toda la información disponible de un periodo de tiempo suficientemente amplio para poder disponer de un alto número de operaciones.

Esto quiere decir que, de las ofertas actuales hay que considerar todas las disponibles, y para las operaciones, todas las que puedan obtenerse que se hayan realizado desde la actualidad hasta 12 ó 24 meses hacia atrás.

Una primera comprobación que hay que realizar es si la muestra seleccionada es representativa, es decir, si para cada variable la distribución de la muestra es similar a la del parque inmobiliario que representa. Esto se lleva a cabo con facilidad mediante gráficos y tablas con porcentajes, que comparan la muestra con la totalidad de propiedades existentes para cada variable. Normalmente cuando la muestra es elevada esto se suele cumplir, pero hay que confirmarlo.

Una vez obtenida la información de mercado en origen, es decir las operaciones de mercado con sus precios y variables, es necesario calificarla.

### **Calificación de la información**

Esta tarea consiste en calificar la información que consideremos válida y descalificar la inválida.

Se trata de analizar la información para detectar operaciones con datos incompletos, erróneos o con alguna variable desproporcionada. Esta es una calificación inicial, pues los algoritmos matemáticos que se usarán más adelante para modelizar el mercado, suelen disponer de herramientas de ayuda que permiten identificar operaciones que pueden distorsionar los resultados, y que habrá que analizar para su posible descalificación.

Para realizar esta tarea se utilizan herramientas estadísticas básicas, tales como gráficos, medias, dispersiones, etc. La descalificación se produce cuando se detectan operaciones que tienen alguna variable incompleta, errónea o fuera de rango. Estas operaciones descalificadas no se deben eliminar, sino que hay que retirarlas a un fichero aparte y argumentar su descalificación. Esto se debe hacer en aras de la transparencia del procedimiento, de manera que las operaciones descalificadas se puedan ver, y se pueda comprobar que el valorador utilizó toda la información disponible, si bien una parte de la información fue descalificada de forma razonada.

Esta tarea se conoce también con el nombre de “limpieza de datos”.

Un aspecto importante del proceso de calificación es el de acotación de las variables. Para cada una de las variables conviene definir los límites máximo y mínimo que puede tener, de modo que las operaciones que excedan estos límites puedan ser descalificadas por atípicas. Por ejemplo, en vivienda colectiva se podría decir que la

superficie de las operaciones válidas debe estar entre 30 y 200 m<sup>2</sup>, de esta forma si tenemos en la muestra un piso de 15 m<sup>2</sup> puede ser descalificado.

Para no hacer esta tarea a sentimiento, conviene ver la distribución de las superficies de todos los pisos existentes, y así elegir estadísticamente los máximos y mínimos de esta variable para la muestra, y lo mismo para todas las variables.

Cuando el modelo matemático este terminado, se usará para valorar todos los productos, incluidos los que excedían los límites. En general los modelos matemáticos suelen tener una cierta capacidad de extrapolación que permite resolver bien la valoración de estos productos atípicos, pero no suele ser conveniente introducirlos para la construcción del modelo, pues por lo general son pocos, anómalos y distorsionantes.

### **Incorporación de variables de situación**

La información en origen suele contener un cierto número de variables relativas a la construcción: antigüedad de la finca, superficies construidas por usos, características de la cubierta, instalaciones, plantas, anejos, etc. Sin embargo, en cuanto a la situación de la propiedad solo suele aportar su posición en el plano y la calle y número correspondientes, y es evidente que estos datos no se pueden procesar así, tal cual. La construcción de un modelo exige que haya también variables que indiquen la mejor o peor situación de los inmuebles.

Estas variables de situación hay que construirlas, y aportárselas a la muestra. Son responsables, en buena parte, del éxito de la valoración. Cuando hayamos generado el modelo y comprobemos que sus resultados no son del todo satisfactorios, las deficiencias serán, en un porcentaje muy alto de los casos, atribuibles a unas variables de situación inadecuadas.

Una forma de obtener estas variables es zonificando el territorio, y obteniendo en cada zona valores medios de la categoría de construcción, de la antigüedad, etc. Cada operación tomará estos datos según la zona en la que se encuentre.

Otra forma consiste en obtener estas mismas variables (categoría media, antigüedad media, etc.), de las fincas próximas, mediante un método como el de K-vecinos, que se explicará más adelante, u otro similar.

También es posible usar distancias a centros de interés.

La zonificación es el sistema más extendido para aportar variables de situación a la muestra. El proceso de comprobación y ajuste del modelo exigirá, a la vista de los defectos observados en los resultados, revisar y corregir la zonificación inicial.

### Tratamiento de la información

Las variables y los precios que forman parte de las operaciones no suelen estar de origen en formatos procesables por un algoritmo matemático. Algunas variables incluso no son numéricas y hay que transformarlas en una secuencia de números procesable. En otras puede ser conveniente agrupar sus datos en escalones (discretización), o transformarlos con funciones diversas. Es preciso pensar para cada variable cual es la forma numérica más adecuada de que entre en el proceso.

Otro tratamiento importante es el de la actualización de los precios.

Cuando se trabaja con operaciones de un determinado periodo de tiempo (por ejemplo de los últimos 12, 18 o 24 meses), es necesario actualizarlos a una fecha para que sean comparables. Hay diversos métodos destinados a obtener la tasa de variación mensual para la actualización de los precios.

Una vez que la información ha sido calificada y tratada es necesario estimar qué variables, de entre todas las disponibles, son las que mejor van a explicar el valor de mercado.

### Selección de variables

La identificación de las variables significativas es una de las tareas más difíciles, y por lo general la realiza el experto de forma intuitiva. Pero esta selección intuitiva suele dar problemas. Una cosa es que pensemos que una variable deba intervenir en la variación del precio, y otra cosa es que al combinarse con las demás variables pueda producir efectos distorsionantes. Las variables interactúan en conjunto y algunas de ellas, que consideradas por separado parecen necesarias, pueden presentar problemas de correlación o deformar la incidencia de otras.

No hay que desechar el conocimiento y la intuición del experto en la selección de variables, pero cuando se manejan grandes ficheros con operaciones de mercado conviene recurrir a técnicas matemáticas que aseguren que la selección de variables es eficaz.

Hay dos tipos de técnicas para analizar las variables:

- Técnicas de eliminación de variables.
- Técnicas de selección de variables.

Las técnicas de eliminación tienen por objeto detectar y suprimir variables que no deben actuar en la construcción del modelo, por ser inoperantes o por estar correlacionadas con otras existentes. Por ejemplo, hay algoritmos para eliminar las variables que tienen un alto porcentaje de valores idénticos, o que la desviación típica de sus valores está por debajo de un determinado umbral. En estos casos la variable se considera poco operativa. También proponen la supresión de variables que tienen un alto grado de correlación con otras.

Las técnicas matemáticas de ayuda a la selección de variables permiten realizar un análisis más completo, orientado a evaluar cómo trabajan en conjunto las variables. Estas técnicas tienen dos componentes: un Buscador, y un Evaluador.

El Buscador selecciona un subconjunto de variables y lo propone para ser evaluado, y el Evaluador mide la capacidad que tiene ese subconjunto para modelizar el resultado, en nuestro caso el valor, y contrasta si es mejor o peor que la de los subconjuntos evaluados anteriormente.

Entre los Buscadores más utilizados cabe mencionar los de búsqueda exhaustiva, los de búsqueda hacia delante, los de búsqueda hacia atrás y los algoritmos genéticos.

La búsqueda exhaustiva consiste en ir seleccionando sistemáticamente todas las posibles combinaciones de variables para ir evaluándolas y verificar así cual es el mejor subconjunto. Esta forma de búsqueda solo puede utilizarse cuando el número de variables disponibles es reducido.

Imaginemos, por ejemplo, que disponemos de un fichero con un elevado número de operaciones de mercado de las que conocemos el precio y 20 variables (la antigüedad, la categoría de construcción, la superficie, etc.). El número de distintos subconjuntos de variables es de  $2^n - 1$  es decir de más de un millón, y si fuesen 30 variables el número de posibles subconjuntos sería superior a los mil millones. Es evidente que resulta prácticamente imposible evaluar uno a uno estos distintos subconjuntos para seleccionar el mejor, por lo que la búsqueda exhaustiva solo se emplea cuando el número de variables es reducido (6 ó 7 como máximo).

La búsqueda hacia delante consiste en seleccionar primero las variables una a una y evaluarlas, la que obtenga el mejor resultado se mantiene, y se añade una segunda de las que quedan. Se van probando todas las que quedan junto con la primera seleccionada y el par que obtenga mejor resultado se mantiene, añadiendo una tercera de las que quedan. El proceso continúa hasta que, añadiendo una nueva variable de las que quedan, los resultados no mejoran.

La búsqueda hacia atrás es similar pero al revés. Primeramente se seleccionan

todas las variables y se evalúan conjuntamente. Después se retira sistemáticamente una variable y se evalúan todos los posibles subconjuntos de  $n-1$  variables, seleccionando el mejor siempre que supere al de considerar todas las variables, y así se van retirando variables hasta que la evaluación no mejora.

El último Buscador que vamos a comentar es el Algoritmo Genético. Es sin duda el más interesante de todos, aunque es más complejo.

Los algoritmos genéticos pertenecen al ámbito de la Inteligencia Artificial. Fueron propuestos por John Holland en 1975, y utilizan las reglas de la genética para resolver problemas matemáticos. Tienen múltiples aplicaciones y una de las más prácticas es la selección de variables.

El esquema de funcionamiento es el siguiente. Supongamos que tenemos 20 variables. El número de posibles subconjuntos de variables es de aproximadamente un millón, por lo que no es viable probarlos todos. Lo que se hace es seleccionar por ejemplo 12 subconjuntos al azar. Cada uno de estos subconjuntos es un "individuo" y los 12 constituyen la "población inicial", además, las variables que forman cada individuo son sus "Genes".

El paso siguiente consiste en evaluar estos 12 subconjuntos (individuos). Los 6 mejores sobreviven, y los 6 peores mueren. Los 6 supervivientes forman tres parejas, es decir que se casan y tienen hijos. Los hijos heredan las características de sus padres. Así pues, un hijo tendrá una parte de sus genes (variables) heredada de uno de sus progenitores y la otra del otro. Pero no todos los genes de un hijo son herencia de sus padres, algunos genes son distintos pues son "mutaciones" y se asignan aleatoriamente.

La nueva población resultante vuelve a evaluarse y sobreviven solo los mejores. El ciclo sigue repitiéndose y las generaciones se van sucediendo. En general se observa que los hijos van teniendo evaluaciones



mejores que las de sus padres, y en unos cuantos ciclos se alcanza un individuo con una evaluación que ya no se mejora. Este individuo es el subconjunto de variables elegido.

Parece increíble que un método de este tipo permita optimizar la selección de variables, pero es que las reglas de la genética funcionan como un reloj para optimizar un resultado, tanto en la naturaleza como en los procesos matemáticos. El emparejamiento, la herencia, la mutación y la selección son reglas que funcionan para conseguir la adaptación a una situación optimizando los resultados. En la naturaleza estos procesos se extienden a lo largo de siglos, pero en un ordenador se pueden emular en pocos segundos.

Se podría pensar que estos algoritmos matemáticos son herramientas complejas que sólo pueden utilizar algunos especialistas, pero no es así. En realidad son rutinas informáticas que se ejecutan poco más que pulsando una tecla, y el usuario no necesita conocer la formulación matemática interna, sino simplemente tener una idea de cómo funcionan y, por supuesto, practicar con ellos.

La mayoría de estos algoritmos, y de los que veremos a continuación, forman parte de programas informáticos de "Minería de Datos" (Data Mining), algunos de los cuales se pueden descargar gratuitamente de Internet.

Hemos visto los principales Buscadores que utilizan las técnicas de selección de variables pero, como ya dijimos, además de un Buscador hace falta un Evaluador que vaya indicando si el subconjunto de variables buscado es mejor o peor que los anteriores.

Los algoritmos de evaluación son de dos tipos: de Filtro, y de Modelo.

Los de Filtro realizan una evaluación somera y rápida, y son muy eficaces, si bien no garantizan una selección óptima de las variables. Un ejemplo de estos es el algoritmo CFS (Hall, M. A. 1998). Este

algoritmo evalúa cada subconjunto de variables, buscando que cada variable separadamente esté correlacionada con el resultado (es decir con el valor), y que, además, las variables no estén correlacionadas entre sí.

Los de tipo Modelo aplican a cada subconjunto de variables el algoritmo matemático que se va a emplear en la siguiente etapa para modelizar el mercado. Este algoritmo suele ser una regresión o una Red Neuronal, por lo que el proceso de evaluación es más costoso pero la selección final de variables es más precisa.

Las técnicas de selección de variables que se han descrito son una ayuda fundamental para el valorador en una tarea que es muy difícil de resolver a mano. Con ellas se evita perder mucho tiempo probando variables a ciegas, e impiden que seleccionemos un juego de variables poco eficaz y mejorable, con el que sólo podemos alcanzar resultados mediocres.

De todas formas, no hay que olvidar que estos algoritmos únicamente ayudan a seleccionar variables de entre las que hemos obtenido, y puede que falte alguna importante que no hemos considerado. El análisis de los resultados del modelo, que se realizará en la etapa tercera, nos indicará la posibilidad de que los resultados insatisfactorios exijan considerar nuevas variables.

## **Normalización de variables**

En muchos casos conviene normalizar las variables antes de procesarlas. Cuando se comparan los valores numéricos de unas variables con los de otras, se observan a menudo cifras de tamaño y recorrido muy distinto, por lo que suele ser conveniente normalizarlas para evitar distorsiones en la creación del modelo matemático.

Hay diversos métodos de normalización, pero es frecuente usar uno de los siguientes.

- Normalización por Mínimo - Máximo.

$$V' = \frac{V - \text{Mín}}{\text{Máx} - \text{Mín}} (\text{nuevo\_Máx} - \text{nuevo\_Mín}) + \text{nuevo\_Mín}$$

Es frecuente utilizar este sistema para que todas las variables vayan de 0 a 1, aunque se puede escoger cualquier otro par de límites.

- Normalización Z.

$$V' = \frac{V - \text{promedio}}{\text{Desv. Estándar}}$$

## Reducción de variables

Hay algoritmos matemáticos que permiten reducir el número de variables significativas. Supongamos que disponemos de una muestra de mercado en la que para cada operación conocemos el precio y 5 variables significativas (antigüedad, superficie, categoría, planta y situación). Estos algoritmos pueden transformar estas 5 variables en 2 variables equivalentes.

Estas variables equivalentes no tienen ya un significado concreto, es decir que no son ni la antigüedad ni la superficie ni nada identificable, pero las podemos llamar X e Y, y podemos dibujar en un gráfico de dispersión todas las operaciones con su valor. Cuando, más adelante, terminemos el proceso de construcción del modelo podremos comparar en un gráfico de este tipo, utilizando las variables reducidas, el valor original con el valor calculado por el modelo y visualizar así la el grado de adecuación de los resultados a los valores originales.

Por tanto, esta reducción a 2 variables equivalentes tiene la utilidad de poder ver en un gráfico de dos dimensiones lo que no se podía ver por estar en N dimensiones (N variables).

La reducción de variables no sólo se puede utilizar para visualización de los datos originales y de los resultados, sino

que también puede aplicarse como proceso previo antes de introducir los datos en el algoritmo matemático que va a generar el modelo de mercado.

Las técnicas de reducción de variables más utilizadas son las siguientes.

- Reducción S.V.D.
- Análisis de Componentes Principales.
- Mapas Auto-organizativos.

La reducción S.V.D. (Single Value Decomposition) está basada en el cálculo matricial.

El Análisis de Componentes Principales reduce las variables preservando la varianza.

Los Mapas Auto-organizativos aplican una red neuronal de dos capas para llevar a cabo la reducción.

Estos algoritmos de reducción de variables, así como los que se han visto anteriormente, son herramientas de pre-procesado de la información que forman parte de los programas informáticos de minería de datos.

Todas las tareas que se han descrito dentro de esta etapa de tratamiento de la información: calificación, transformación, actualización de precios, selección de variables, normalización y reducción, son procesos aplicables a la información en origen como preparación antes de introducirla en un algoritmo matemático de aprendizaje que permita obtener el modelo de mercado.

En general, suelen ser todas ellas necesarias para poder alcanzar un resultado aceptable. Es interesante conocer y experimentar los algoritmos matemáticos que las soportan para identificar cuales son los más adecuados en cada caso.

De todo lo dicho se desprende que esta etapa es muy laboriosa, y que requiere bastante tiempo, de hecho consume la mayor parte del tiempo y del esfuerzo necesario para la construcción del modelo. De todas formas es una etapa muy interesante, pues en ella hay que tomar muchas decisiones, hacer pruebas, y analizar los datos y las variables con detalle, todo lo cual propor-

ciona al valorador un conocimiento profundo de la información de mercado que tiene entre manos.

Una vez completada esta etapa disponemos de la información de mercado preparada para ser procesada por un algoritmo matemático de aprendizaje.

## Modelo matemático

En esta segunda etapa se construye el modelo de mercado utilizando un procedimiento matemático estándar.

Esta etapa solo tendrá éxito si la etapa anterior ha sido minuciosa y sistemática. Hay que tener en cuenta que la aplicación de los algoritmos matemáticos de aprendizaje es bastante sencilla, y en muchas ocasiones apenas lleva unos minutos obtener los resultados, pero enseguida habrá que comprobarlos, y estudiar qué ajustes hay que hacer, pues es muy difícil que estén bien en el primer intento.

Hay diversos tipos de algoritmos matemáticos capaces de modelizar los valores de mercado. Lo que hacen, en general, es leer y procesar el fichero con las operaciones de mercado depuradas y tratadas en la etapa anterior, aplicando una rutina capaz de aprender y de extraer la relación entre las variables y el valor.

Entre los procedimientos más empleados para modelizar el valor de mercado de los productos inmobiliarios están los siguientes.

- Regresiones.
- Redes Neuronales.
- K-Vecinos.

### Regresiones

La regresión múltiple es un procedimiento muy conocido, que está implementado en programas informáticos de uso común, tales como las hojas de cálculo.

Es la expresión más inmediata del Método Hedónico, y en ella se trata de poner el valor en función de las variables mediante una ecuación sencilla, de tipo lineal, exponencial, cuadrática, etc.

Las modalidades más utilizadas para el mercado inmobiliario son las siguientes.

- Regresión aditiva.
- Regresión multiplicativa.
- Regresión híbrida.

En la primera, la ecuación resultante es del tipo:

$$Y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n$$

Donde “Y” es el Valor, “ $x_i$ ” las variables y “ $a_i$ ” los coeficientes establecidos por la regresión.

La regresión multiplicativa tiene dos modalidades:

$$Y = a_0 * a_1^{x_1} * a_2^{x_2} * \dots * a_n^{x_n}$$

$$Y = a_0 * x_1^{a_1} * x_2^{a_2} * \dots * x_n^{a_n}$$

La primera de estas dos modalidades es la que suele aportar mejores resultados, y es la más utilizada.

En cuanto a la regresión híbrida, se llama así por ser una mezcla de las dos anteriores, y su ecuación es del tipo:

$$Y = a_1^{x_1} * a_2^{x_2} * \dots + \dots * a_{n-1}^{x_{n-1}} * a_n^{x_n}$$

Es decir que tiene dos partes multiplicativas separadas por una suma.

La regresión aditiva relaciona las variables con el valor mediante la ecuación de una línea recta en el espacio de “n” dimensiones, siendo “n” el número de variables. La contribución de cada variable al valor es bastante clara al añadirse mediante una suma a las demás.

La regresión multiplicativa hace lo mismo que la anterior, pero esta vez la ecuación corresponde a una línea curva

de tipo exponencial. Este tipo de regresión permite obtener, por lo general, resultados más precisos que el anterior. Esto es así porque la relación de las variables con el valor no tiene por que aproximarse a una línea recta, y suele tener alguna curvatura.

La regresión híbrida recoge ventajas de las dos anteriores. Busca la mayor precisión de la multiplicativa, combinada con la claridad de la contribución de las variables de la aditiva, si bien aquí son dos grupos de variables. Suele ser frecuente agrupar en uno de los sumandos las variables cuantitativas y en el otro las cualitativas.

Cualquiera que sea el tipo de regresión que se utilice, su aplicación al fichero con la muestra de mercado produce el resultado inmediatamente. El resultado es la ecuación, junto con una serie de indicadores que sirven para estimar su calidad.

Lo normal es que la ecuación de regresión no salga del todo bien en un primer intento, y los indicadores muestran como mejorarla. Una de las mejores indicaciones es la identificación de "outliers", que son operaciones atípicas, o con datos erróneos que superaron la primera etapa. Al retirar estas operaciones y volver a ejecutar la regresión los resultados mejoran significativamente.

La regresión se muestra, en general, como un sistema eficaz para construir un modelo matemático del mercado, y son muchos los profesionales que la utilizan para realizar las valoraciones administrativas en diversos Estados. En un primer momento fue la regresión de tipo aditivo la que comenzó a usarse, pero en la actualidad las de tipo multiplicativo e híbrido están ganando terreno por su mayor precisión.

Sin embargo, la regresión presenta algunas limitaciones que conviene apuntar. En primer lugar no es muy tolerante a fallos, es decir que unas pocas operaciones erróneas o atípicas pueden alterar mucho los resultados. Esto puede ser un problema en

un mercado como el inmobiliario donde, como ya se ha comentado, la información presenta bastantes deficiencias.

Además, tiene problemas para modelizar el mercado cuando las variables se relacionan con el valor de una forma compleja, es decir que la relación no se adapta a una línea recta o a una curva sencilla. Por ello suele dar buenos resultados cuando se aplica a una urbanización o a una ciudad pequeña o mediana. Pero cuando el territorio aumenta, y con él la complejidad de las variables, los resultados pueden deteriorarse.

También presenta deficiencias para valorar propiedades atípicas.

De todas formas, conociendo estas limitaciones y seleccionando adecuadamente las variables y las operaciones, este método es eficaz como lo muestra la extensión de su uso.

## Redes Neuronales

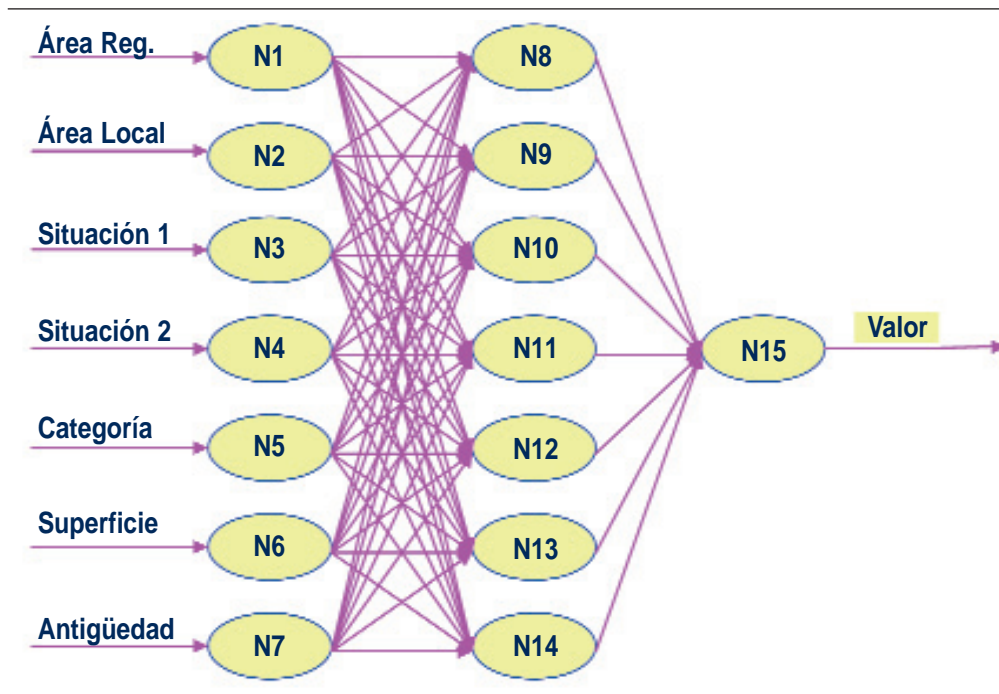
Las Redes Neuronales son una modalidad de la Inteligencia Artificial basada en el estudio del sistema nervioso y de sus células, las neuronas.

Los algoritmos matemáticos de este tipo emulan una red de neuronas preparada para leer el fichero con la muestra de mercado y aprender la relación entre las variables y los precios.

Las neuronas artificiales no ajustan la relación entre las variables y el precio a un determinado tipo de línea, como sucede en la regresión, sino que mediante un proceso de aprendizaje consistente en probar una solución y después corregir el error, se adaptan a la relación existente entre las variables y el precio, por lo que son más flexibles. De hecho en este proceso utilizan funciones no lineales, como la función logística.

Las Redes Neuronales que se utilizan para la modelización del mercado inmobiliario son de tipo Feed-Forward con tres capas según el esquema indicado en la Figura 2.

Figura 2  
Red Neuronal de tres capas



La primera capa tiene una neurona por cada variable, y se limita a recibir los datos de las operaciones y transmitirlos tal cual a la siguiente, llamada capa oculta o intermedia. Esta segunda capa procesa la información y la transmite a la última capa, que solo tiene una neurona y que genera el valor de mercado.

El proceso es cíclico. Las operaciones entran varias veces en el sistema, y para cada una de ellas se estima el valor, se obtiene el error como diferencia entre el valor estimado y el real, y se procede a corregir este error propagándolo hacia atrás para ajustar todas las neuronas. Este proceso se llama aprendizaje o entrenamiento de la Red. El cálculo se detiene cuando la reducción del error apenas progresa, y en un punto en el que se evite el sobre-entrenamiento.

Una vez entrenada la Red Neuronal, esa es la “fórmula” que permitirá valorar los

inmuebles. Es una fórmula compleja pero que se ha construido automáticamente y que se aplica con facilidad.

Este modelo de tres capas que hemos comentado se conoce también con el nombre de “Perceptrón”.

Existen Redes Neuronales más complejas, con más capas y con un diseño de conexiones y fórmulas internas diferentes, adecuadas para modelizar el mercado usando variables territoriales continuas. El tipo de Red conocido como Mapa Auto-asociativo, que se mencionó en el apartado de Reducción de Variables, también se ha utilizado para valorar. Pero la Red de tres capas descrita es, con mucho, el tipo más utilizado y produce unos resultados muy satisfactorios.

Las redes neuronales tienen ventajas e inconvenientes, respecto a la regresión, que conviene precisar.

Entre los inconvenientes está el hecho de que el conocimiento técnico para el manejo de las Redes es algo superior, pues el valorador tiene que ajustar algunos parámetros tales como la tasa de aprendizaje, el momento, y los ciclos. De todas formas el manejo de Las Redes a nivel de usuario no es complicado, y con un pequeño curso y algo de experiencia se adquiere cierta confianza.

Otra desventaja es que los cálculos internos que hace el sistema son complejos y difíciles de exponer, lo que puede generar desconfianza en las personas que reciban los resultados. La regresión también presenta cierta complejidad, pero no tanta. Aquí conviene apuntar que el mercado inmobiliario es un asunto difícil, que la relación de las variables con el precio es compleja, y que si se busca mayor precisión, normalmente será a costa de la sencillez expositiva.

De todas formas, hay que considerar que, las Redes, al igual que las regresiones o cualquier otro algoritmo matemático, son herramientas de trabajo del experto. Una vez que haya alcanzado los resultados, lo normal es que los refleje en tablas y planos con zonificaciones que le permitan exponer sus conclusiones con facilidad a los no iniciados.

Entre las ventajas que tienen las Redes Neuronales respecto a las regresiones hay que citar la tolerancia a fallos. Ya comentamos que las regresiones presentan una baja tolerancia a fallos, sin embargo las Redes tienen una tolerancia bastante alta, y son capaces de manejar ficheros de datos que contengan un cierto porcentaje de información errónea sin que los resultados se vean afectados de forma relevante.

Esta capacidad de trabajar con información de baja calidad es muy importante en nuestro caso pues, como ya se ha comentado, la información del mercado inmobiliario presenta deficiencias importantes, por la propia naturaleza del mercado, que no se solucionan del todo en la etapa primera de tratamiento de la información.

Otra ventaja de las Redes Neuronales respecto a las regresiones es la mayor precisión

que alcanzan en sus resultados. Cuando el objeto a modelizar es sencillo, una urbanización, o un municipio pequeño o mediano, los resultados son de una precisión similar en ambas técnicas. Pero cuando el objeto es más complejo y el territorio aumenta de tamaño, las Redes muestran una mayor precisión. Por esto pueden enfrentarse a la valoración de grandes territorios manteniendo una precisión aceptable y consiguiendo un nivel de producción de valores muy elevado.

La última ventaja que conviene citar es su capacidad para aproximar mejor el valor de las propiedades especiales o atípicas.

Estas dos últimas ventajas proceden del hecho de que las Redes no ajustan la relación entre las variables y el precio a una ecuación sencilla, como en la regresión, sino que aprenden una relación más compleja pues sus funciones no lineales las dotan de mayor flexibilidad.

Veamos un ejemplo. Supongamos que disponemos de 17 operaciones de mercado relativas a pisos de las mismas características y en la misma ubicación, que solo varían entre sí en la superficie. Los pisos son de 200, 190, 180, ..., 40 y 30 m<sup>2</sup>. Como los pisos son iguales, lo único que hace variar el valor es la superficie, por tanto, para estos ejemplos, el valor solo está en función de la superficie.

En la Figura 3 vemos el gráfico con la serie de valores real correspondiente a la muestra, y la serie de valores resultante de aplicar una regresión aditiva a los datos reales.

La serie de valores unitarios reales es curva ya que el valor unitario va aumentando cada vez más pronunciadamente a medida que el tamaño del piso disminuye. Sin embargo la regresión aditiva es una línea recta. Es la mejor línea recta que se puede encajar en esa curva, pero como vemos el error que se comete al sustituir la curva real por la línea de regresión es de cierta importancia.

En el gráfico de la Figura 4 añadimos la línea correspondiente a una regresión multiplicativa.

Figura 3. Regresión aditiva (Valor según el tamaño de la vivienda)

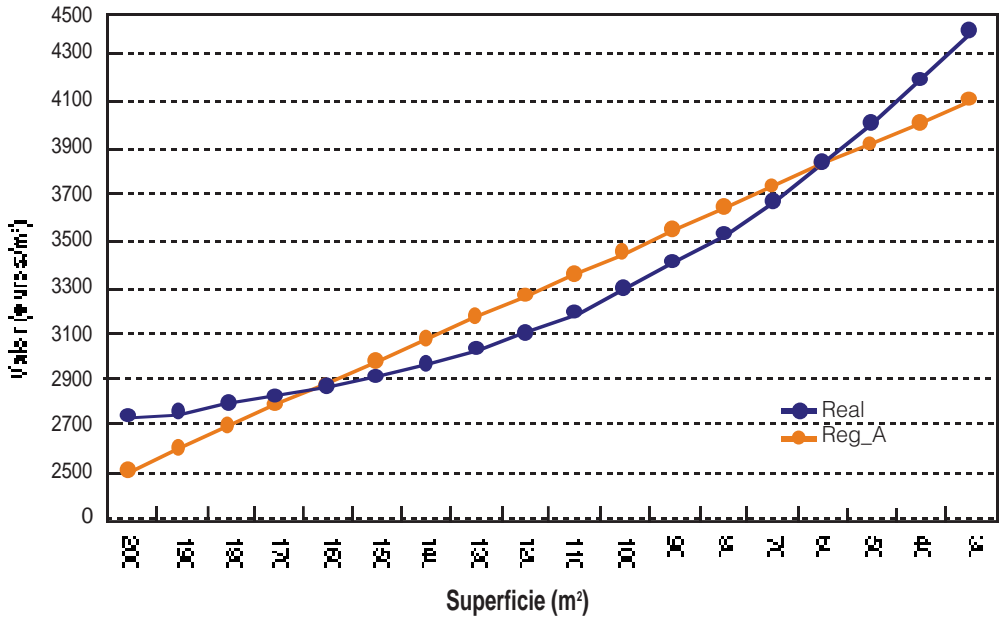


Figura 4. Regresión aditiva y multiplicativa (Valor según el tamaño de la vivienda)

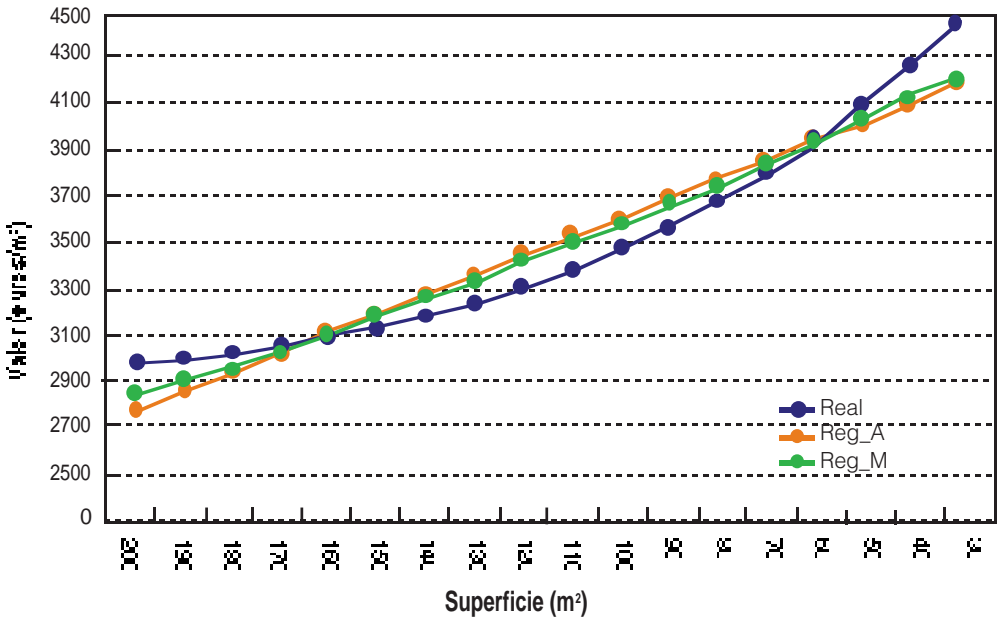
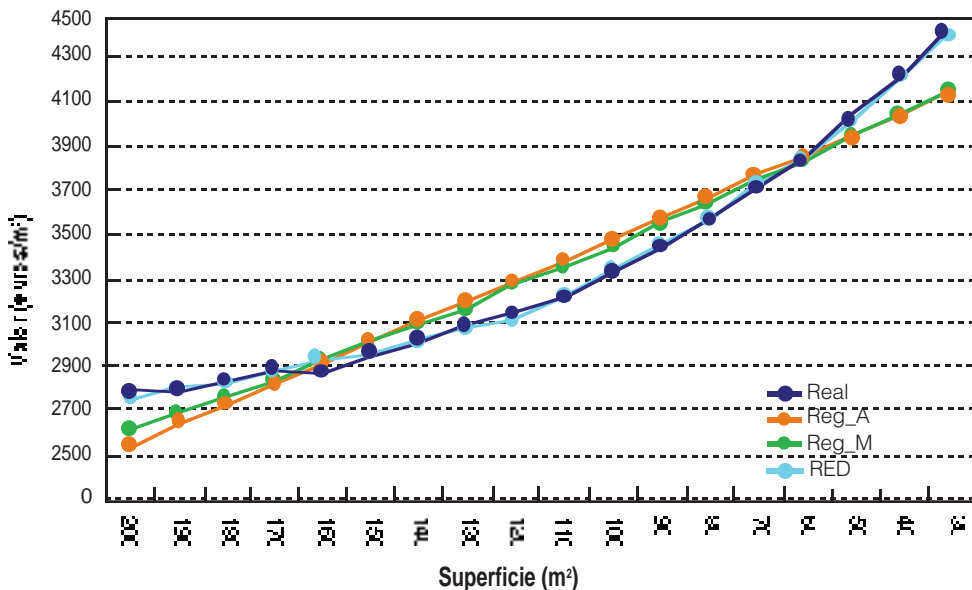


Figura 5  
 Regresiones y Red Neuronal (Valor según el tamaño de la vivienda)



La nueva línea mejora un poco a la anterior pero sigue cometiendo un cierto error porque, aunque es una curva exponencial, la curva de los valores reales no es una curva exponencial.

Por último, en el gráfico de la Figura 5, añadimos la línea resultante de modelizar los valores reales con una Red Neuronal.

El resultado es casi idéntico a la curva real, lo que muestra la flexibilidad de los Modelos Neuronales, que no se ajustan a un tipo de ecuación predeterminada sino que se adaptan a la curvatura contenida en la muestra.

Continuando con las comparaciones, hay que decir que una regresión es básicamente una Red Neuronal sin capa intermedia, tal y como se indica en el esquema de la Figura 6.

Según los expertos, esta forma de operar relacionando las variables directamente con el precio mediante un simple coeficiente es demasiado tosca. La relación entre las variables y el precio es suficientemente

compleja como para que no pueda alcanzarse con precisión comunicándolos directamente.

Hace falta una capa intermedia de procesadores que elabore la información de entrada y la prepare para alcanzar el resultado.

En el gráfico de la Figura 7 se compara el error de la regresión con el de la Red Neuronal para un fichero con operaciones de vivienda colectiva.

El error de la regresión es el mismo desde el primer momento, pero el de la Red disminuye a medida que avanzan los ciclos de aprendizaje. A los 10 ciclos la Red alcanza el mismo error que la regresión, pero tiene capacidad para seguir mejorando y elaborando un modelo más preciso a medida que los ciclos de entrenamiento avanzan.

En la Gerencia Regional del Catastro de Madrid se han utilizado las Redes Neuronales como herramienta de trabajo para valorar las 2.100.000 viviendas colectivas que hay



Figura 6  
Esquema del modelo de Regresión

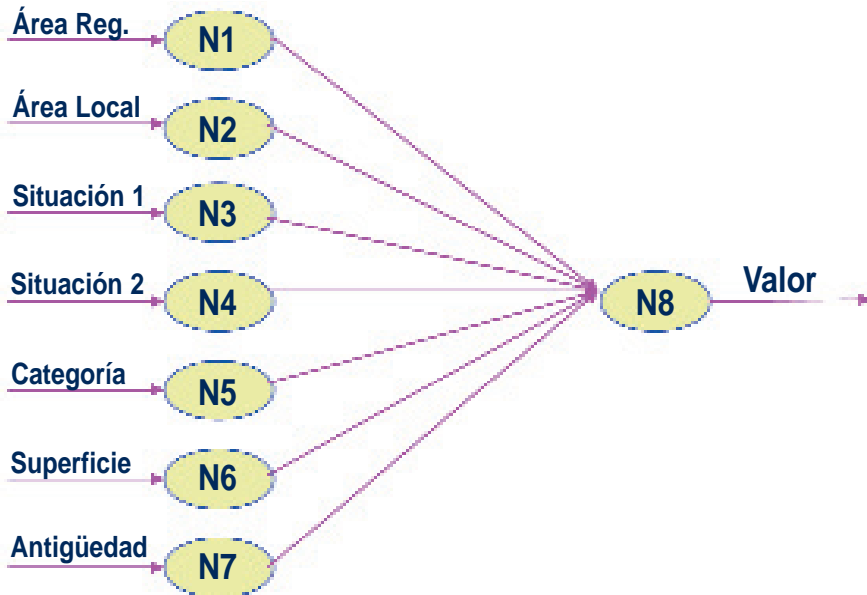
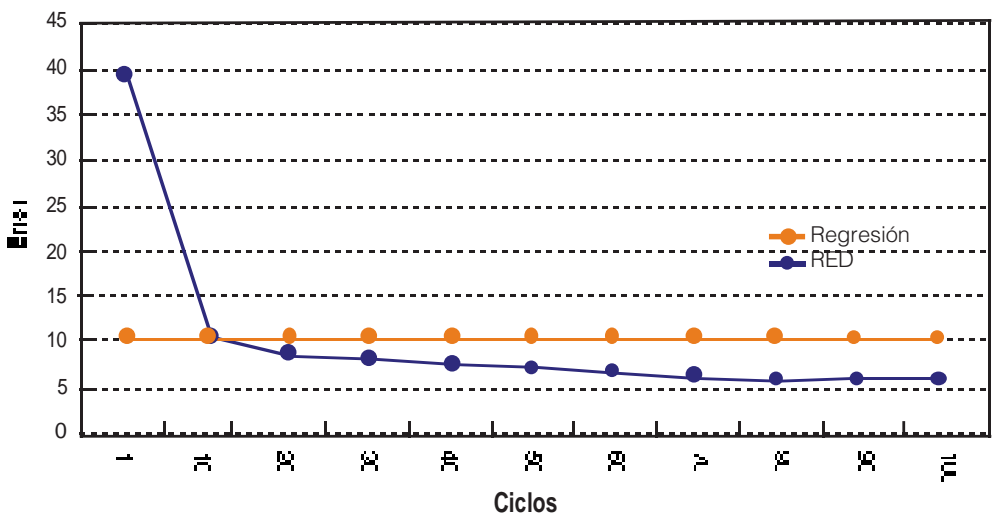


Figura 7  
Evolución del error (Aprendizaje de la Red Neuronal)



en la Comunidad de Madrid, como prueba para confirmar las capacidades de esta tecnología. Las valoraciones se han llevado a cabo de forma completa en los años 2005, 2006 y 2007, con resultados aceptables, y también se ha realizado el modelo para la valoración de la vivienda unifamiliar, que se aplicará en su caso a 400.000 propiedades.

Estos trabajos de investigación muestran la gran capacidad de los modelos de valoración automatizada para la producción de valores, manteniendo buenos niveles de precisión.

## K-Vecinos

El algoritmo de K-Vecinos (K-Nearest-Neighbors, Fix y Hodges 1951) se diferencia de los anteriores en que no genera ninguna fórmula. La valoración de un inmueble se lleva a cabo con los inmuebles más similares a él, que existan en el fichero de operaciones, es decir con sus vecinos más próximos.

Se parte de un fichero de operaciones u ofertas de las que conocemos su precio, y al que hemos aplicado todas las tareas correspondientes a la etapa primera: calificación, tratamiento, selección de variables, etc.

Para valorar un inmueble, el primer paso consiste en determinar cuales son sus vecinos más próximos del fichero que contiene las operaciones de las que conocemos el precio.

Para determinar la mayor o menor proximidad se recurre a calcular la distancia entre el inmueble que queremos valorar y los del fichero con la muestra. Pero la distancia que se calcula no es una distancia física, sino lógica. Es decir que en ella intervienen todas las variables, no solo las de situación, sino también las de superficie, antigüedad, etc.

Por tanto es necesario calcular la distancia entre la operación a valorar, con variables  $(x_{01}, x_{02}, x_{03}, \dots, x_{0n})$ , y cada una de las operaciones del fichero de muestra, con variables  $(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in})$ .

Existen diferentes métodos para calcular la distancia entre dos operaciones, pero el más utilizado es la distancia euclídea. De todas formas los programas que soportan este algoritmo dan a elegir entre varios tipos de distancias: Euclídea, Manhattan, Mahalanobis, Minkowski, Chevishev, etc.

Una vez que se dispone de la distancia lógica entre el inmueble a valorar y todos los de precio conocido, se seleccionan los más próximos, es decir los de distancia menor.

Para seleccionar los vecinos más próximos hay que decidir primero qué número de vecinos va a intervenir en el cálculo. Según elijamos un número de vecinos u otro el resultado variará, por lo que es conveniente tomar esta decisión de modo fundamentado. Un número de vecinos pequeño hará que el valor sea más sensible a los precios erróneos que pudiera haber, y un número de vecinos grande puede hacernos perder precisión.

En los esquemas de la Figura 8 vemos dos ejemplos con la influencia del número de vecinos en el resultado.

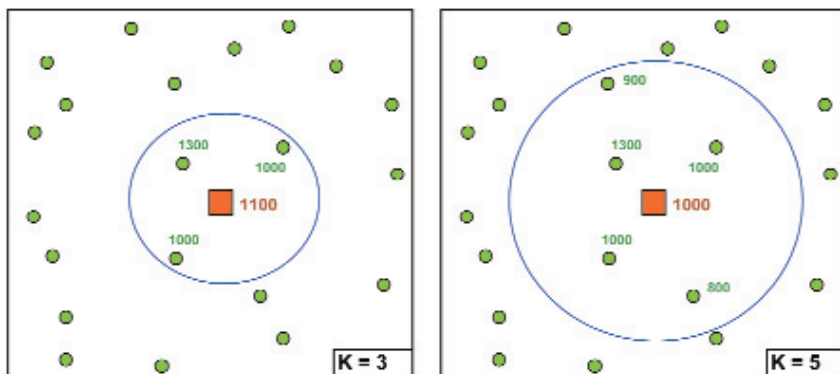
La forma de elegir el número de vecinos (K) consiste en aplicar el cálculo del valor a las operaciones del fichero de muestra, de las cuales conocemos el precio. Hacemos el cálculo para  $K = 2, 3, 4, 5, \dots, n$  sucesivamente, y obtenemos el error para cada uno de estos K como diferencia entre los valores obtenidos y los precios reales. El K para el que se obtenga un menor error con el fichero de muestra es el elegido.

Una vez elegido el número de vecinos, ya podemos valorar cualquier inmueble. El algoritmo calcula su distancia respecto a todas las operaciones del fichero de muestra, selecciona los K vecinos más próximos, y con ellos determina el valor.

Para determinar el valor con los K vecinos más próximos hay tres posibilidades entre las que el usuario debe elegir.

- Valor medio
- Valor ponderado con la distancia
- Método Global

Figura 8  
Método de K-vecinos



En el primero el valor se calcula como promedio del valor de los K vecinos más próximos.

En el segundo cada uno de los K vecinos realiza una aportación al valor inversa al cuadrado de la distancia, de modo que los más alejados influyen menos.

El método global es similar al anterior, solo que aquí intervienen todas las operaciones (no solo las K más próximas), con aportaciones al valor inversas al cuadrado de la distancia. Las operaciones que estén muy alejadas influirán muy poco.

Una aplicación adecuada de este algoritmo exige escalar las variables, para que no haya ninguna que influya mucho más que las demás en el cálculo de la distancia. Se puede ver fácilmente, que el algoritmo de K vecinos no es más que una formulación matemática del procedimiento tradicional que utilizan los tasadores para valorar los inmuebles apoyándose en testigos próximos.

La ventaja fundamental que presenta este algoritmo frente a los anteriores es que es muy intuitivo y fácilmente explicable a personas no iniciadas en la valoración. Carece de fórmulas complejas y puede alcanzar resultados bastante precisos. El inconveniente es que necesita bastantes

operaciones de mercado repartidas según todo el recorrido de las variables. Si hay escasez de operaciones en alguna zona determinada, los resultados se deformarán.

Existen otros algoritmos que pueden aplicarse a la valoración inmobiliaria, y seguramente que en los próximos años veremos avances en este sentido de las técnicas de Minería de Datos e Inteligencia Artificial. Pero, de momento, los mencionados son los más conocidos.

Una vez que se ha obtenido el Modelo hay que comprobarlo.

## Comprobación y Ajuste del Modelo

Por mucho cuidado que se haya tenido en la selección de la muestra y en la ejecución del algoritmo, no se puede dar por bueno el modelo resultante hasta que no se hagan las comprobaciones oportunas.

El método de trabajo ayuda, pero no garantiza la precisión de los resultados. Con el mismo método se puede hacer una valoración buena o mala. En las valoraciones lo importante no es el método sino acertar.

Hay dos tipos de comprobaciones que hay que realizar.

- Interna: con las propias operaciones que han servido para crear el modelo.
- Externa: con otras operaciones u ofertas externas.

En ambas hay que medir la precisión de la valoración calculando el error, es decir la diferencia entre el precio conocido y el valor calculado.

Hay diversas formas de medir el error: error absoluto, error relativo, precisión, etc. Pero lo importante es conocer si el error que hemos cometido es aceptable, es decir si se encuentra dentro de unos límites establecidos.

En este sentido cabe mencionar los Estándares sobre Ratios de la Asociación Internacional de Valoradores (I.A.A.O.) International Association of Assessing Officers). Veamos un breve apunte sobre estos estándares.

Su mérito consiste en establecer unas condiciones mínimas que debe cumplir la valoración para poder considerarla correcta. Estas condiciones se refieren a medidas estadísticas sencillas, con límites numéricos. Otra ventaja es que tienen un alto grado de aceptación en muchos países.

Lo primero que hay que hacer es calcular el Ratio para todas las operaciones (tanto de la comprobación interna como de la externa),

$$\text{Ratio} = \text{Valor Calculado} * 100 / \text{Precio.}$$

Por tanto, el Ratio es el porcentaje del valor calculado respecto al precio, esto es, que si el Ratio es un 100% es por que el valor calculado y el precio son iguales.

Después se toman tres medidas.

- Medida de Centralidad: se calcula la Mediana del Ratio de todas las operaciones. Esta medida sirve para com-

probar si la valoración está, en conjunto, alta, baja o ajustada.

- Medida de Dispersión: para esta medida se utiliza el Coeficiente de dispersión que se expresa en porcentaje. Una dispersión baja indica una mayor precisión de la valoración.
- Diferencial de precios: Indica si la valoración es progresiva o regresiva, es decir si se han valorado más bajo las propiedades más caras o las más baratas.

El cumplimiento simultáneo de estos tres estándares asegura un nivel de precisión aceptable para la valoración masiva. Hay que señalar, no obstante, que los estándares se deben cumplir en el conjunto del territorio a valorar, y también en cada una de sus partes: barrios, zonas, etc. Es frecuente en los primeros intentos, que los estándares se cumplan globalmente, pero que haya un 20 o un 30% de zonas donde no se cumplan, y sea necesario revisar las variables de situación y las operaciones para ajustar el modelo.

Para la comprobación con datos de mercado externos, que no han intervenido en el proceso de construcción del modelo, hay que tomar las mismas precauciones que para las operaciones que sí han intervenido. Es decir que todas las tareas de la etapa primera, de tratamiento de la información, son aplicables a este fichero de muestra externa, con los mismos criterios (para variables, para normalización, etc.) que para la muestra interna.

Hay que prestar especial atención a la distribución de la muestra según las variables, para que sea lo más representativa posible, ya que con ella vamos a ajustar el modelo.

Lo normal es que, al probar el modelo por primera vez con una muestra externa los resultados no sean satisfactorios. Esto no debe desanimarnos. Hay que analizar los resultados y detectar donde puede estar el problema. A veces el problema se cir-

cunscribe a determinadas zonas en las que hay que revisar los datos de las operaciones o el contenido de las variables de situación. Otras veces se detectan anomalías con alguna variable, que no funciona como se esperaba. En cualquier caso hay que indagar con detalle las causas del error.

Una vez detectadas y corregidas las causas del error volvemos a la primera etapa, y repetimos el proceso de obtención del modelo. Es un trabajo de ajuste cíclico, con el que vamos afinando los resultados.

Para realizar este trabajo, los algoritmos de modelización suelen contar con herramientas que facilitan el análisis para detectar el origen de los errores. Estas herramientas permiten, por ejemplo, estudiar el comportamiento de las variables, localizar el error, o identificar operaciones que están teniendo un efecto distorsionante en el modelo.

Una de las indicaciones de estas herramientas, es la descalificación de operaciones de mercado que habían superado la primera etapa de calificación. En cualquier caso, estas operaciones no hay que descalificarlas sin más, sino que, al igual que en la primera etapa hay que analizar por qué no funcionan bien y argumentar su descalificación.

Hay muchos motivos diferentes por los que un modelo puede fallar: operaciones, variables, normalización, parámetros de los algoritmos, etc. Por ello la etapa de ajuste no es fácil, pero en la práctica los problemas se acaban detectando. En esto se notan mucho las horas de vuelo, por lo que no conviene abordar modelos muy complejos para empezar. Tampoco hay que entusiasmarse con los primeros resultados, sino tener espíritu crítico y analizarlos con detalle.

Construir un Modelo de Valoración Automatizada, como se ha visto, es muy laborioso. Hay que desterrar la idea de que solo consiste aplicar un algoritmo a un fichero de muestra, ya que con esa idea no se llega a ninguna aparte. Pero cuando se

ha conseguido, es relativamente sencillo mantenerlo en el tiempo, o hacer otros modelos similares y, además, la capacidad de producción de valores que se consigue es superior a la de los métodos tradicionales.

Una vez que el modelo está ajustado ya solo falta utilizarlo para valorar.

## Aplicación del Modelo

Las aplicaciones más frecuentes de un Modelo de Valoración Automatizada son las siguientes.

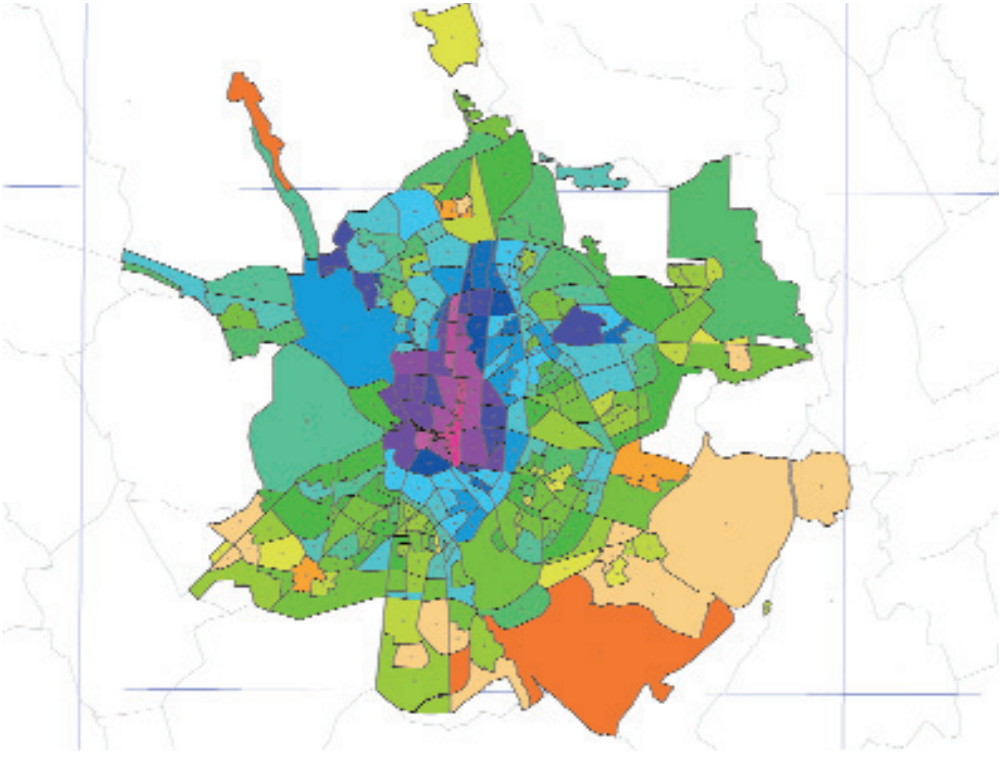
- Valoración Individual.
- Valoración Masiva.
- Mapas de Valores.

El modelo puede utilizarse como una “Calculadora de Valores”, de modo que se le introducen los datos de un determinado inmueble y se obtiene el valor. Esta utilidad puede ser interesante para ponerla al servicio de usuarios, que no necesitan saber nada del modelo, y que lo único que quieren es introducir los datos y ver el resultado.

La Valoración Masiva es otra de sus aplicaciones más interesantes. Lo que hay que hacer es preparar todos los inmuebles a valorar, con el mismo formato de la muestra, y aplicar el modelo. Los algoritmos descritos suelen valorar bases de datos enormes en cuestión de segundos.

Por último están los Mapas de Valores. Estos mapas son una tradición para los técnicos que hacen valoración masiva. El modelo permite valorar un inmueble tipo en cada zona del territorio. Si después descontamos el coste de la construcción de ese inmueble y obtenemos el valor del suelo, podemos representarlo en un plano. La figura 9 contiene un mapa de valor de suelo del municipio de Madrid realizado con una Red Neuronal.

Figura 9  
Mapa de valores del suelo de Madrid



Estos mapas son de gran ayuda en el proceso de ajuste del modelo, pues permiten ver los gradientes de valor y sus posibles inconsistencias.

En definitiva, hay que decir que hay tecnología cada vez más avanzada que puede ser aplicada a la valoración, que hay información de mercado disponible cada vez más abundante, que es difícil sacar todo el partido a esta información sin ayuda de herramientas matemáticas, y que conviene que los valoradores al menos conozcamos la existencia y el funcionamiento de estos sistemas en la medida en que pueden ir cambiando nuestra forma de trabajar.

### **Bibliografía**

- JOHN HOLLAND. *Adaptation in Natural and Artificial Systems* (1975).
- International Association of Assessing Officers. *Standards on Ratio Studies* (y otros estándares).
- E. FIX, J. L. HODGES J.R. *Discriminatory analysis, non parametric discrimination* (1951).
- M. A. HALL. *Correlation-based feature selection for discrete and numeric class machine learning* (2000).
- VOLKER STRUMPEN, HENRY HOFFMAN, AND ANANT AGARWOL. *A stream algorithm for the s.v.d.*
- TEUVO KOHONEN. *Intro to Self Organizing Maps* (2006).
- EERO CARLSON. *Kohonen Map, GIS and the Analysis of the Real Estate Sales*, FIG International Congress (2002). ■