

Las Redes Neuronales en Investigación de Mercados

Héctor Tavera

Gallup México

En muchas ocasiones hemos escuchado que *debemos aprender de nuestras experiencias*. Este adagio es especialmente útil en el mundo de los negocios, donde lo principal es, a veces, predecir el futuro sobre la base del aprendizaje que se ha obtenido. La tecnología de las redes neuronales es una herramienta que está emergiendo en: investigación de mercados, análisis financiero y consultoría de los negocios. Se está utilizando para predecir las tendencias futuras con base en las pasadas. En el mundo de la investigación de mercados, las redes neuronales pueden ser usadas, entre muchas otras cosas, para responder a preguntas tales como:

- ◆ ¿Qué tanto se recordará mi marca en el próximo trimestre?
- ◆ ¿Por qué mis consumidores son más proclives a responder por correo directo?
- ◆ ¿Cómo puedo seleccionar la mejor ubicación para mis nuevas sucursales?

Estos tipos de preguntas han sido difíciles de responder, y en algunos casos hasta imposible, con las tradicionales herramientas estadísticas. En este artículo se explicará qué son las redes neuronales y haremos una pequeña comparación con dos herramientas multivariadas tradicionales.

¿Qué Son las Redes Neuronales?

En el nivel más básico, las redes neuronales son simplemente un nuevo camino para el análisis de los datos. A pesar de que ahora se cuenta con herramientas tecnológicas muy avanzadas respecto a las redes neuronales, el origen de esta herramienta se remonta a los 40's. Este concepto está basado en sistemas computarizados que son capaces de comprender los patrones de los datos obtenidos anteriormente. Igual que cada ser humano aprende cómo hablar, caminar y distinguir objetos a través de la experiencia, las redes neuronales aprenden de los datos pasados para reconocer los patrones existentes.

Las redes neuronales complementan los análisis estadísticos tradicionales. A diferencia de éstos, las redes neuronales no requieren suposiciones sobre la forma de la distribución de los datos para poder ser analizados (linealidad, normalidad, etcétera), supuestos que muy pocas veces se cumplen en la investigación de mercados.

Aplicación: Clasificación de Cartera

Como bien sabemos, quienes nos dedicamos a la investigación de mercados, existen varias técnicas estadísticas multivariadas que nos permiten la segmentación para usarla posteriormente como método de clasificación.

Se ejemplificará la clasificación de la cartera de un cliente con base en los datos de investigaciones anteriores, de manera que pueda clasificar a sus futuros clientes.

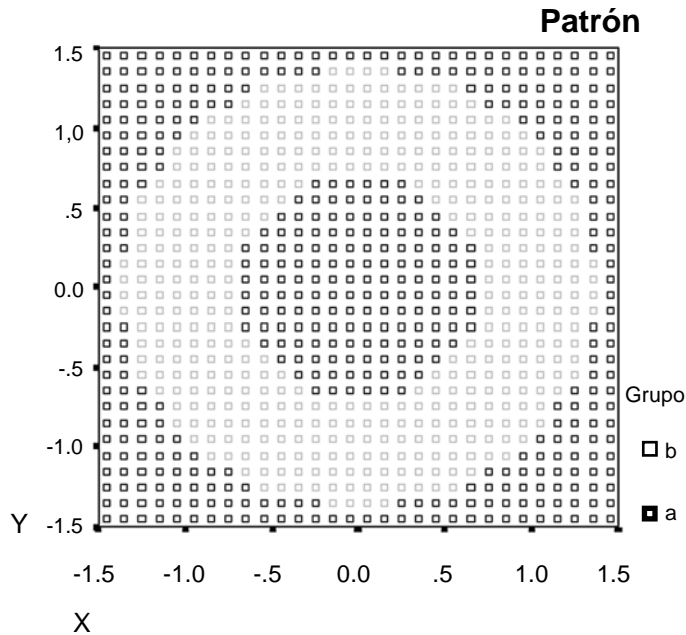
Para dicha clasificación se mostrarán los resultados más importantes de tres técnicas estadísticas utilizadas:

- ◆ Análisis Discriminante.
- ◆ Algoritmo de Classification & Regression Tree (Arboles de clasificación y regresión).
- ◆ Redes Neuronales.

Primeramente observemos el patrón de los datos ($n=10,000$), el cual está constituido por dos grupos. Dicho patrón como se puede observar, es de forma circular.

Este es un problema interesante cuando se pretende clasificar un conjunto de datos, porque:

- ◆ No hay una separación en las medias de los grupos; esto es, las medias en X y Y son idénticas en ambos casos.
- ◆ La frontera entre los dos grupos no es lineal y de hecho son una curva continua.



Para poder comparar las tres técnicas estadísticas se tomó una muestra aleatoria (n=800) de los datos originales para validar el modelo. De esta manera tenemos una muestra generadora de la segmentación, y otra para la clasificación de los datos en las tres técnicas. Como la intención es la clasificación correcta de la muestra aleatoria, se presentarán los resultados de clasificación. Cabe aclarar que el rigor estadístico en cada una de las técnicas fue validado correctamente y por falta de espacio no se presentarán en esta ocasión.

Análisis Discriminante

Se realizó un análisis discriminante simple (variable dependiente formada por dos grupos), utilizando la función discriminante lineal de Fisher. En la tabla 1 se muestran los resultados de la clasificación del análisis discriminante. En ésta se observa que sólomente el 51.1% de los datos de la muestra aleatoria son clasificados correctamente; siendo ésta una clasificación muy pobre.

Tabla 1

Resultados de la clasificación b,c

	SEG	Grupo de pertenencia		Total	
		0	1		
Original	Recuento	0	197	387	
		1	201	413	
	%	0	50.9	49.1	100.0
		1	48.7	51.3	100.0
Validación cruzada	Recuento	0	183	387	
		1	223	413	
	%	0	47.3	52.7	100.0
		1	54.0	46.0	100.0

- La validación cruzada sólo se aplica a los casos del análisis. En la validación cruzada, cada caso se clasifica mediante las funciones derivadas a partir del resto de los casos.
- Clasificados correctamente el 51.1% de los casos agrupados originales.
- Clasificados correctamente el 46.6% de los casos agrupados, validados mediante validación cruzada.

Classification & Regression Tree (C&RT)

Ahora, en la tabla 2, se pueden observar los resultados arrojados por el análisis de los datos utilizando C&RT. Debe citarse que ambos árboles, tanto el de construcción como el de clasificación, fueron desarrollados usando el método de “desarrollar el árbol y podar” y con un mínimo número de casos en el nodo parental de 25 y en el nodo filial de 10.

Tabla 2

Matriz de clasificación errónea				
		Categoría real		
		a	b	Total
Categoría estimada	a	183	15	198
	b	230	372	602
	Total	413	387	800
		Reestimación		
Estimación de riesgo		0.30625		
ET de la estimación de riesgo		0.0162965		

El porcentaje de clasificación errónea es del 30% (estimación de riesgo) y se puede notar que una gran cantidad del grupo “a” (categoría real) es mal clasificado en el grupo “b” (categoría estimada).

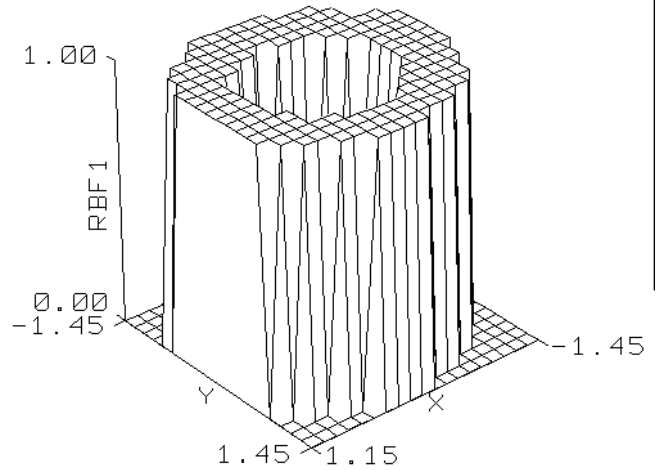
Redes Neuronales

Las redes neuronales generan un modelo aprendiendo el patrón de los datos, al no forzar una ecuación global para todos los datos. Otra de las fortalezas es que no son necesarias las suposiciones sobre los datos o asumir modelos y probar después su capacidad de clasificación. Por último, aplicando esta técnica a los datos de la muestra aleatoria se observan los siguientes resultados.

Se observa que las redes neuronales clasifican erróneamente el 3.1% (Percentage Incorrect) de los datos de clasificación. Aún más, si se observa el gráfi



Field number:	<input type="text" value="5"/>	Target bin ranges:									
Field Name:	<input type="text" value="Output 1"/>	a									
Number of bins:	<input type="text" value="2"/>	b									
Percentage Correct:	<input type="text" value="96.900002"/>										
Percentage Incorrect:	<input type="text" value="3.1"/>										
Number of records processed:	<input type="text" value="800"/>	Actual bin ranges:									
RMS Error:	<input type="text" value="0.000000"/>	a									
Percentage out of range:	<input type="text" value="0.0"/>	b									
Mean Absolute Error:	<input type="text" value="0.000000"/>										
Cross tabulation matrix:	Actual										
Target	<table border="1"> <tr> <td></td> <td>a</td> <td>b</td> </tr> <tr> <td>a</td> <td>386</td> <td>11</td> </tr> <tr> <td>b</td> <td>14</td> <td>390</td> </tr> </table>			a	b	a	386	11	b	14	390
	a	b									
a	386	11									
b	14	390									



co, se puede notar un patrón muy similar al de los datos originales. Esto nos lleva a sugerir que los datos de los estudios anteriores del cliente son piezas clave para descubrir el patrón de los datos.

Muchas organizaciones en el mundo se benefician de la habilidad de predecir el futuro, mediante la combinación de la tecnología de las redes neuronales y los datos de estudios anteriores. Esta capacidad de predecir es accesible en nuestros días. Nuestra experiencia pasada es más que historia: ésta contiene información valiosa la cual nos abre una ventana al futuro.

¿ QUIERES TENER EL MEJOR SOFTWARE PARA INVESTIGACIÓN DE MERCADOS ?

Ventas y Soporte Técnico
 (5)117 14 50
 Pino 307, Col. Santa María Insurgentes.
 C.P. 06430, México D.F.

Visite nuestra página web
www.cfmc.com



SURVENT
 El Sistema CATI más poderoso.

WEB-SURVENT
 CATI en Internet.

SOUND - SURVENT
 Obtenga la idea real a través del sonido real.

MENTOR
 Sistema de tabulación para profesionales en Investigación de Mercado.

COSI
 Programa de tabulación y gráficas en Windows.

AHORA EN MÉXICO