

Miguel Estevez (2017-0200)

Resumen:

Uso de redes neuronales y múltiple regresión lineal para predecir el rendimiento de los bancos en Malasia.

Puntos Importantes:

- La data proviene de 13 bancos de Malasia entre los años 2001-2006 elegidos aleatoriamente de la base de datos BANKSCOPE.
- El rendimiento de un banco se puede representar con el retorno en activos (el que se usó en este paper), retorno en capital y el retorno en depósitos.
- En todos los estudios anteriores que comparan los métodos tradicionales estadísticos con la inteligencia artificial, esta última le gana a todos.
- Variables a utilizar para calcular el ROA(retorno en activos):
 - LIQ = proporción de préstamo-activos como medida de la liquidez
 - LLOS = proporción de pérdida de provisión con préstamo como medida de riesgo de crédito
 - COSTINC = relación costo-ingreso
 - $SIZE_i$ = Tamaño del banco basado en su total de activos
 - $CONC_i$ = proporción de concentración, calculado al coger los 3 más grandes bancos dividido por el total de activos
 - CPI_i = Índice de precios al consumidor
 - GDP_i = grosor de producto doméstico
 - ϵ_i = termino de error
- De los 96 data que fueron coleccionados (6 años X 13 bancos seleccionados), 86 se usaron para el training set y 10 para el testing set.
- La función de activación que se utilizó es la función sigmoid :

$$\phi(v) = \frac{1}{1 + e^{-v}}$$

Resultados de la regresion lineal multiple:

Al coger las 7 variables, tabla 2.

Model Summary						
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson	
1	0.670	0.449	0.442	0.86267		
2	0.756	0.572	0.561	0.76496	0.991	
ANOVA Results						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	50.891	1	50.891	68.383	.000
	Residual	62.513	84	0.744		
	Total	113.403	85			
2	Regression	64.835	2	32.417	55.399	.000
	Residual	48.569	83	0.585		
	Total	113.403	85			
Coefficients						
Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.873	0.326		11.867	0.000
	<i>COSTINC</i>	-0.064	0.008	-0.670	-8.269	0.000
	(Constant)	3.881	0.289		13.410	0.000
2	<i>COSTINC</i>	-0.061	0.007	-0.637	-8.834	0.000
	<i>LLOSS</i>	-0.199	0.041	-0.352	-4.881	0.000

- Las variables más significativas eran: riesgo en crédito y relacion costo-ingreso.

$$B_0 = 3.881, b_2 = -0.199 \text{ y } B_4 = -0.061$$

La ecuacion de regresion lineal multiple:

$$ROA_i = B_0 + B_1 LIQ_i + B_2 LLOS_i + B_3 SIZE_i + B_4 COSTINC_i + B_5 COC_i + B_6 GDP_i + B_7 CPI_i + \epsilon_i$$

$i = 1, \dots, n$

La funcion regresion robusta es $3.423 - 0.4197LLOSS - 0.0288COSTINC$

Resultado de la red neuronales:

Se utilizó una red de multicapa de perceptrón con una capa oculta y 7 entradas, que corresponden a las 7 variables del estudio. Aquí se compara los resultados cambiando el número de neuronas donde el 13 es el mejor.

Neurons	Training Data		Test data	
	MSE	R ²	MSE	R ²
5	0.02040	0.81364	0.17329	0.85239
10	0.00799	0.94691	0.01500	0.83672
11	0.00505	0.96707	0.02444	0.65187
12	0.00515	0.96438	0.20735	0.59641
13	0.00358	0.97767	0.00687	0.66868
14	0.00529	0.96644	0.00964	0.79567
15	0.00297	0.96683	0.04524	0.60258
20	0.00428	0.97182	0.00738	0.87044
25	0.00672	0.96680	0.03358	0.64716

Para comparar ambos resultados se utilizó la media de error predecido donde la red neuronal gana por mucho a la regresión múltiple, como se ve a continuación:

Method	MSPR
Multiple Linear Regression	0.6190
Artificial Neural Network	0.0061

Opinion

A mi parecer este paper es muy interesante ya que demuestra cómo las redes neuronales pueden implementarse en muchas áreas, en este caso las finanzas y ganarle a las técnicas que se llevan utilizando desde hace un tiempo en esas áreas. Lo cual podría llevar a pensar en otras áreas de conocimiento donde se pudiera implementar estos algoritmos para optimizar los resultados. Una de las desventajas que son mencionadas en esta investigación es que vieron que para encontrar el número óptimo de neuronas se toma cierto tiempo ya que se deben hacer varias iteraciones del mismo algoritmo para encontrarlo. Pero esto también puede ser por los equipos que se utilizaron ya que esta investigación es del 2009 y el panorama de la computación ahora es muy diferente a lo que había en esos tiempos.