

Ejemplo de Aplicación con *Neural Connection*

El marketing *directo* es un área donde las redes neuronales han tenido un considerable éxito.

En este tipo de marketing, se vende un producto o un servicio, enviando una oferta directamente al cliente: no hay ni tiendas, ni intermediarios entre la compañía que vende el producto y el cliente que lo compra.

Aunque esta forma puede dar altos beneficios está sometida también a mayores riesgos, de manera que es conveniente controlar o clasificar adecuadamente a los clientes potenciales.

Por medio de las redes neuronales se pueden orientar los objetivos básicos de las ofertas, y se pueden analizar las relaciones complejas entre los atributos de los consumidores y las posibles ofertas.

El ejemplo que se considera es un problema que se le presenta a una compañía que vende mercancías a sus clientes por marketing directo y está preparándose para introducir un nuevo producto. La compañía quiere saber cuáles de sus clientes habituales responderían positivamente a la nueva oferta. Así, la compañía espera eliminar el costo del envío, sin perder negocio.

Para resolver este problema, se puede construir una aplicación que predice si un cliente entra en una de dos categorías: los que responderán a la oferta, y los que no. Además, se da una medida del éxito, en términos del porcentaje de clientes correctamente clasificados.

Se tienen los siguientes datos:

VAR1	VAR2	VAR3	VAR4	VAR5	VAR6
56.0	Guildford	90.62	1.0	90.69568	non-respond
28.0	Glasgow	80.0	1.0	90.69568	respond
38.0	London	200.0	3.0	90.6446	non-respond
35.0	Cardiff	80.0	1.0	90.69568	respond
38.0	Glasgow	320.0	1.0	90.6446	respond
.....					
44.0	Guildford	230.0	2.0	90.69568	non-respond
77.0	Bristol	80.0	1.0	90.63923	respond
74.0	Guildford	349.40	4.0	90.69568	non-respond
65.0	Guildford	80.0	1.0	90.78976	respond
50.0	Cambridge	140.0	1.0	90.63923	non-respond
43.0	London	140.0	1.0	91.44568	non-respond

La compañía había lanzado un producto similar un año antes. Para esa oferta, la compañía registró los clientes que respondieron, y los clientes que no. Incluyendo la información que tuvieron sobre su base de datos de clientes, la compañía pudo disponer de un conjunto de datos con la información siguiente para cada cliente:

- Edad
- La oficina regional más cercana
- El número de pedidos hecho en el último trimestre
- La cantidad de dinero gastada en pedidos en el año pasado
- La fecha en la que primero compraron mediante la compañía

- La respuesta al lanzamiento de un producto similar

Se pretende construir un modelo que asocie la respuesta de un cliente con la edad del cliente, región, número de pedidos, gasto anual total, y fecha de primer contacto con la compañía. La red neuronal en su aplicación construirá un modelo de estas asociaciones que permitirá predecir la respuesta de clientes por adelantado. Las aportaciones a la red neuronal serán las cinco variables con información anterior, y el rendimiento será la respuesta del cliente.

Se resuelve el problema con NETAGENT:

1. Empezar NEURAL CONNECTION.
2. Clicar en **NetAgent** (al lado de **file**) y clicar en **run**.
3. En la ventana de diálogo, seleccionar el fichero **net.agt** de la carpeta **samples**.
4. Empezar el NETAGENT clicando en **Play** (en forma de triángulo en la esquina izquierda) hasta que NETAGENT pregunte si se quiere hacer en el modo **Beginner** o en el modo **Advanced**.
5. Clicar en **Beginner**.
6. Clicar en **Play**.
7. Clicar en **Classification**. Se trata de decidir si un caso (esto es, un cliente potencial) pertenece a la clase de *responders*, o a la clase de *no-responders*.
8. Clicar en **Play**.
9. Usar la lista desplegada y seleccionar **Neural**.

10. Clicar en **Play**. Se usará una función de base radial. Esta es una red neuronal supervisada, **feed-forward** con una capa oculta de neuronas artificiales. Utiliza básicamente funciones no lineales.
11. Clicar en **Play**, hasta que se pregunte el nombre del fichero de datos.
12. Escribir el nombre del fichero, `tutor.sav`.
13. Clicar en **Play**.
14. Usar la lista desplegada y seleccionar **Test data**. Se quiere medir si es adecuado el modelo para resolver el problema.
15. Clicar en **Play**.
16. Clicar en **Play**, hasta que el *training* empiece.

Una vez que el *training* haya terminado, el NETAGENT habrá terminado, quedándose el programa en la pantalla de entrada (WORKSPACE), pudiéndose ver los resultados de la aplicación.

Para ver los resultados:

En la imagen **Results** pinchar encima y seleccionar **View**. Ir a la última columna donde se encuentran las posibles respuestas: *respon* o *non-resp*.

Otro ejemplo

En el enfoque anterior se miraba si un caso (un cliente potencial) pertenecía a la clase de *responders*, o a la clase de *no-responders*. Un enfoque mejor para la compañía sería recodificar el problema para determinar la probabilidad de que un cliente responda positivamente.

Esto resulta ser mucho más útil: la base de datos de clientes podría organizarse ordenándose según la probabilidad de que se responda. Se podrían considerar un número de estrategias diferentes para encontrar el escenario de máximo beneficio para la compañía.

Se subdivide el conjunto de datos en dos secciones diferentes:

- Datos recogidos durante la campaña previa, para usarlos como *training* y *testing*.
- Datos procedentes de la base de datos de clientes, para ser clasificados y usados en la nueva campaña.

Los datos para *training* y *testing* están en el fichero `tutor.sav`.

La base de datos de clientes para clasificar están en el fichero `tutor_rn.sav`.

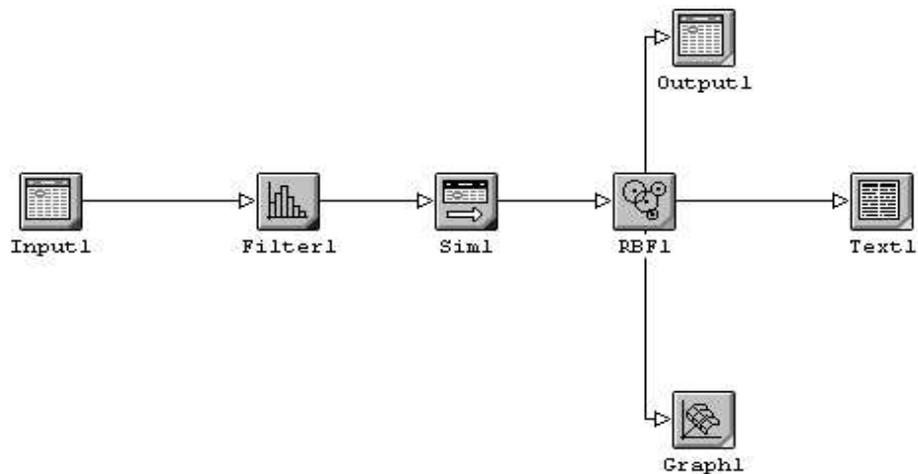
Este tipo de problema, donde se quiere relacionar un conjunto de *inputs* (comportamiento de clientes) respecto a un *output* (¿responderán los clientes?), se debe resolver usando una red neuronal supervisada.

En el programa se encuentran tres tipos de redes supervisadas: con función de base radial, redes neuronales bayesianas y perceptrón multicapa. En este caso se va a usar la función de base radial.

El diseño que se va a emplear consiste en:

- DATA INPUT para importar los datos

- **FILTRO** para usarlo durante la experimentación
- **SIMULATOR TOOL** que permite representar las gráficas.
- **RADIAL BASIS FUNCTION**, la técnica de red neuronal para construir el modelo.
- **TEXT OUTPUT** para examinar los resultados.
- **DATA OUTPUT** para sacar los resultados.
- **GRAPHICS OUTPUT** para examinar el modelo de red neuronal.



Una vez completado el diseño se podrá construir la aplicación y empezar el proceso de *training* y *testing*.

1. Seleccionar **New** del menú **File**.

2. Seleccionar el DATA INPUT TOOL (primera imagen del grupo) en la TOOLS PALETTE, clicando encima.
3. Mover el DATA INPUT TOOL hacia el lado izquierdo con el ratón y fijarlo allí clicando con el ratón.

En general, para mover o colocar cualquier elemento en la pantalla de trabajo, basta con:

1. Clicar encima con el ratón.
2. Mantener apretado el botón izquierdo del ratón.
3. Mover la imagen a la posición deseada con el ratón y soltar después.

A continuación se tienen que introducir los datos de entrenamiento para analizar el problema. Están almacenados en el fichero `tutor.sav`.

Para cargar los datos abrir el DATA INPUT TOOL, y decir que abra el `tutor.sav`.

1. Clicar una vez en el DATA INPUT TOOL con el cursor. Aparece entonces el menú.
2. Del menú DATA INPUT TOOL, elegir **View**. Aparece, entonces una hoja en blanco.
3. En el menú **File**, elegir **Open**. Aparece el diálogo de **Data Input**. Aquí se puede especificar el fichero, que sirve para entrenar y testear la aplicación. El fichero de trabajo se usará cuando el modelo esté implementado. Ambos ficheros van por separado.
4. Seleccionar el botón de **Flat-File radio**.

5. Clicar en **Configure**. Vamos a introducir, en lo que sigue, el fichero que se va a utilizar en la aplicación.
6. En la carpeta: **list box**, elegir la carpeta **samples**.
7. En la lista de tipos de ficheros, elegir **SPSS (*.sav)**.
8. Elegir, entonces, **tutor.sav**.
9. Clicar en **OK**.
10. En el diálogo DATA INPUT, clicar **Close**.

Los datos ahora se cargarán en el **tool** de DATA INPUT. En el **Data Input tool** se asume que la fila primera de datos son los nombres de la columna. El **Data Input tool** asume, también, que el último campo cargado es el campo que contiene los datos *output*, y que los otros campos son los *inputs*.

El DATA INPUT TOOL automáticamente selecciona qué elementos de los datos se usarán para el *training*, la validación, y el testeo del modelo.

En el fichero **tutor.sav**, el *output* se ha codificado como un campo simbólico con dos valores: *respond* y *non-resp*. Como se quiere predecir una probabilidad, que es un número, hay que cambiar a un campo numérico.

1. Ir a la última columna y clicar en el campo denominado **Symbol**. Se abre, entonces, el diálogo de conversión.
2. Seleccionar **Float** de la lista correspondiente a **Current**. El símbolo de *non-resp* se convierte en 0,000000
3. Seleccionar **respond** en la lista de **From Values/Ranges**.

4. Cambiar el valor en el **Convert To**: 1,000000
5. Clicar en **Replace**.
6. Clicar en **OK** para cerrar el cuadro de diálogo.

A continuación se construye el resto de la aplicación.

Desde el submenú **File** del DATA VIEWER , elegir **Exit&Return**.

Se va a usar un FILTER, un SIMULATOR, una red neuronal Radial BASIS FUNCTION, una GRAPHICS OUTPUT TOOL una DATA OUTPUT TOOL y una TEXT OUTPUT TOOL.

1. Colocar un FILTER TOOL en la página a la derecha del DATA INPUT TOOL.
2. Colocar una SIMULATOR TOOL en la página a la derecha del FILTER TOOL.
3. Colocar una RADIAL BASIS FUNCTION (RBF1) en la página a la derecha del SIMULATOR TOOL.
4. Colocar una TEXT OUTPUT TOOL en la página a la derecha y encima de la RADIAL BASIS FUNCTION.
5. Colocar una DATA OUTPUT TOOL en la página a la derecha de la RADIAL BASIS FUNCTION, y directamente debajo de la TEXT OUTPUT TOOL.
6. Colocar una GRAPHICS OUTPUT TOOL en la página a la derecha de y debajo de la RADIAL BASIS FUNCTION, y directamente debajo del DATA OUTPUT TOOL.

Se conectan ahora todas las herramientas:

1. Clicar en el DATA INPUT TOOL.

2. Desde el menú de DATA INPUT TOOL, elegir **Connect**.
3. Clicar en el FILTER TOOL. Se establecen entonces conexiones entre el DATA INPUT TOOL y el FILTER TOOL.
4. De manera similar hay que seguir los pasos anteriores para conectar el resto de herramientas:
 - El FILTER al SIMULATOR.
 - El SIMULATOR a la RADIAL BASIS FUNCTION.
 - La RADIAL BASIS FUNCTION al TEXT OUTPUT.
 - La RADIAL BASIS FUNCTION al DATA OUTPUT.
 - La RADIAL BASIS FUNCTION al GRAPHICS OUTPUT.

NOTA: En caso de rectificación, se pueden eliminar las conexiones seleccionando **Disconnect** del menú de herramientas.

Antes de correr la aplicación hay desde el TEXT OUTPUT hay que cambiar el formato en el que se producen los resultados. Se obtiene una matriz y el porcentaje correctamente clasificado.

En este ejemplo, el indicador del éxito es decidir de modo correcto si un cliente va a responder o no a la campaña. Los valores de los estadísticos en el TEXT OUTPUT se pueden alterar para indicar esto, usando una matriz de cruce con dos valores, uno entre 0.0 y 0.5 (que responde probablemente), y otro entre 0.5 y 1.0 (que no responde probablemente).

Para cambiar los estadísticos en el TEXT OUTPUT:

1. Clicar en el **TEXT OUTPUT**.
2. Desde el menú **TEXT OUTPUT**, elegir **select Dialog**.
3. En el **DATA OUTPUT DIALOG**, clicar **Format**.
4. En el cuadro de diálogo de **OUTPUT FORMAT**, escribir un 2 en el campo correspondiente a **Bins**.
5. En el cuadro de diálogo de **OUTPUT FORMAT**, clicar **OK**.
6. Clicar en **OK**.

Se corre ahora la aplicación. Se seleccionan automáticamente los parámetros para la red neuronal. Mientras se entrena la **RADIAL BASIS FUNCTION**, aparece el cuadro de diálogo de **TRAINING DIALOG** y luego los resultados.

Desde el menú **TEXT OUTPUT**, elegir **Run**.

Ver los resultados y abandonar la pantalla:

- En la barra de menú, clicar **Text Display Off**.

Se encuentran buenos resultados que se pueden mejorar usando un preprocesamiento de los datos: Usando el **FILTER TOOL** para analizar y ponderar los datos.

Se observa que la tercera variable (la cantidad de dinero gastado en el último año) es muy asimétrica hacia la derecha.

Para analizar la tercera variable:

1. Clicar en el **FILTER TOOL**.
2. Desde el **FILTER MENU**, elegir **Dialog**.
3. Clicar en **Var_0003**.

4. Clicar en **Analyze**.

Los valores se pueden ponderar para hacer la distribución más simétrica (así más próxima a la normal). Esto facilita la resolución del problema por la función radial. Se usa entonces la transformación logaritmo:

1. En la ventana de **Analysis**, clicar **Exit**.
2. Clicar en la función de ponderación en la casilla para **Var_0003**.
3. Clicar en $\ln(x + a)$.
4. Reanalizar clicando **Analyze**.
5. Cerrar el **FILTER TOOL**: En la ventana de **Analysis**, clicar **Exit**. En el cuadro de diálogo de **Filter** clicar **OK**.

Volver a correr la aplicación:

- Desde el menú de **TEXT OUTPUT**, elegir **Run**.

Ver los resultados y abandonar la pantalla:

- En la barra de menú, clicar **Text Display Off**.

Examinar las relaciones entre *inputs* y *outputs* de la red neuronal: Se dibujan las variables *input*: customer age, y number of purchases in the past year frente al *output* producido por la red. Se hace con el **SIMULATOR**.

1. Clicar en en el **SIMULATOR**.
2. Desde el menú de **SIMULATOR**, elegir **Dialog**.
3. En **Var_0001**, fijar el **Scan Order** a 1 y **Num Values** a 21.

4. En **Var_0004**, fijar el **Scan Order** a 2 y **Num Values** a 21.
5. Seleccionar **Simulator Enabled**.
6. Clicar en **OK**.

Se produce un gráfico de **Var_0001** (customer age) y **Var_0004** (number of purchases in the past year) frente a la verosimilitud de responder.

- Desde el menú de **GRAPHICS OUTPUT**, elegir **Run**.

Después de ver el gráfico, clicar **Graphics Display Off**

1. Clicar en la **SIMULATOR TOOL**.
2. Desde el menú de **SIMULATOR**, elegir **Dialog**.
3. Deseleccionar **Simulator Enabled**.
4. Clicar en **OK**.

Implementación

Después de examinar los gráficos empieza la fase de implementación. Se pasa la base de datos de los clientes por la aplicación, y se genera un *output* que se usa para ordenarlos según su verosimilitud de respuesta.

Se carga el fichero **tutor_rn.sav** en el **Data Input**.

1. Clicar en en el **Data Input**.
2. Desde el menú de **DATA INPUT**, elegir **View**.
3. Desde el menú **File**, elegir **Open**. Aparece la ventana **Data Input**.

4. En el grupo de **Run Data Input**, clicar **Flat-File**.
5. Clicar en **Configure**. Aparece una ventana **File Open**.
6. Elegir **samples**
7. Elegir tipo **SPSS (*.sav)**.
8. Elegir **tutor_rn.sav**.
9. Clicar en **OK**.
10. En **Data Input**, clicar **Close**.
11. Cerrar el **Data Viewer**: elegir **Exit & Return**.

Se usa la herramienta DATA OUTPUT.

1. Clicar en en Data OUTPUT TOOL.
2. Elegir **View**. Aparece una página en blanco.
3. En **File**, elegir **Setup**. Ahora se puede especificar dónde están los datos.
4. Clicar en **Flat-File** en **Run Data Output**.
5. Clicar en **Configure**. Se abre un **Save** como ventana.
6. De las carpetas elegir **samples folder**.
7. Elegir **SPSS (*.sav)**.
8. En la casilla **File Name**, escribir **results.sav**.
9. Clicar en **OK**.

10. Clicar en **Close**.
11. Desde el menú **View**, elegir **Run**.
12. Desde el menú de **File**, elegir **Output Run Data**.

Los resultados se graban como `results.sav`, y se presentan en la pantalla.

La compañía podría, entonces, ordenar a sus clientes de acuerdo a los *outputs* producidos por la red neuronal.

Práctica con redes Bayesianas:

1. Ir a **File**
2. Ir a **Open**
3. Abrir en **Samples** el fichero: `bayesian.nni`
4. Abrir el **Input1**

Los datos se refieren a varias medidas realizadas en una serie de plantas del género Iris, con vistas a poder hacer una clasificación automática en las tres diferentes especies (*setosa*, *versicolor* y *virginica*), dependiendo de las distintas medidas.

5. Clicar en en el icono de **Bayes1** y luego en **dialog**
6. Observar las opciones por defecto
7. Abrir icono de **Output1**. Observar la ultima columna.
8. Ir a la paleta de iconos y seleccionar un icono de **Text1**
9. Conectar al icono de **Bayes1**

10. Volver a entrenar la red
11. Abrir el icono de **Text1** (clicando en **dialog**)
12. Observar los resultados

Práctica con varios métodos:

1. Ir a **File**
2. Ir a **Open**
3. Abrir en **Samples** el fichero: `bayesian.nni`
4. Observar las salidas de los tres métodos empleados

Práctica con redes de *Kohonen*:

1. Ir a **File**
2. Ir a **Open**
3. Abrir en **Samples** el fichero: `weld.nni`
4. Observar las salida de **Text1**.