

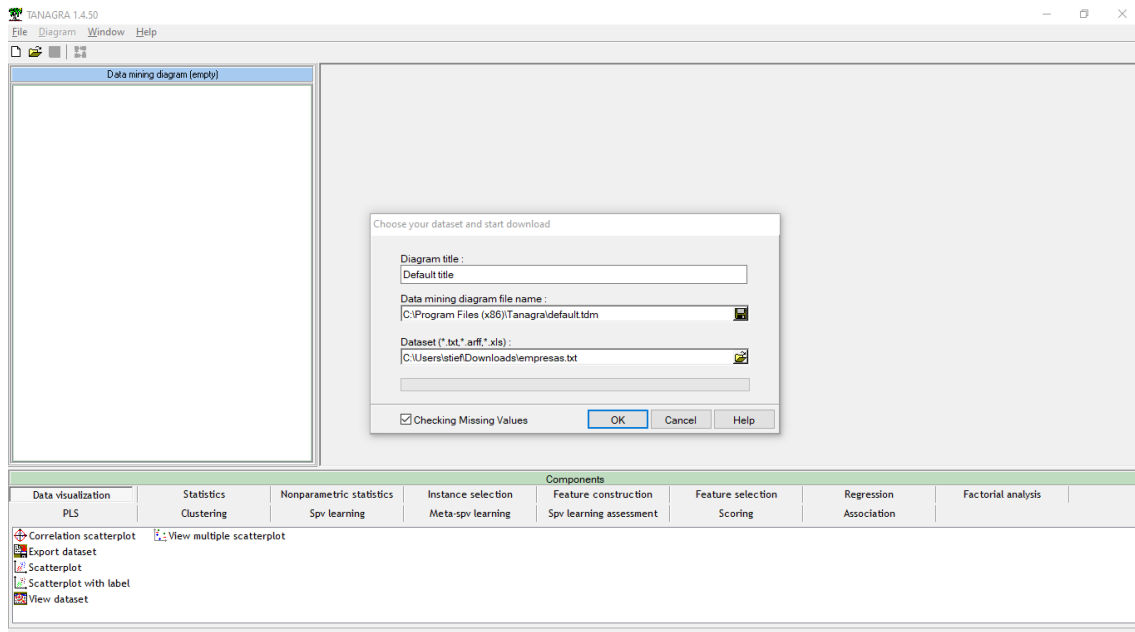
19 Mayo 2020. Tarea Support Vector Machines. Tanagra. Parte 1

Archivo: Empresas.txt

1. Descargar y cargar el archivo de datos.
2. Define status: input: las cuatro variables cuantitativas. Target: Grupo
3. Spv- Learning, SVM (sin cambiar los parámetros por defecto). (observe: exponente=1)
4. Analice la matriz de confusión
5. Indique el peso de cada atributo de entrada y el término independiente
6. Debajo de SVM cuelgue View Dataset. Qué representa la última columna?

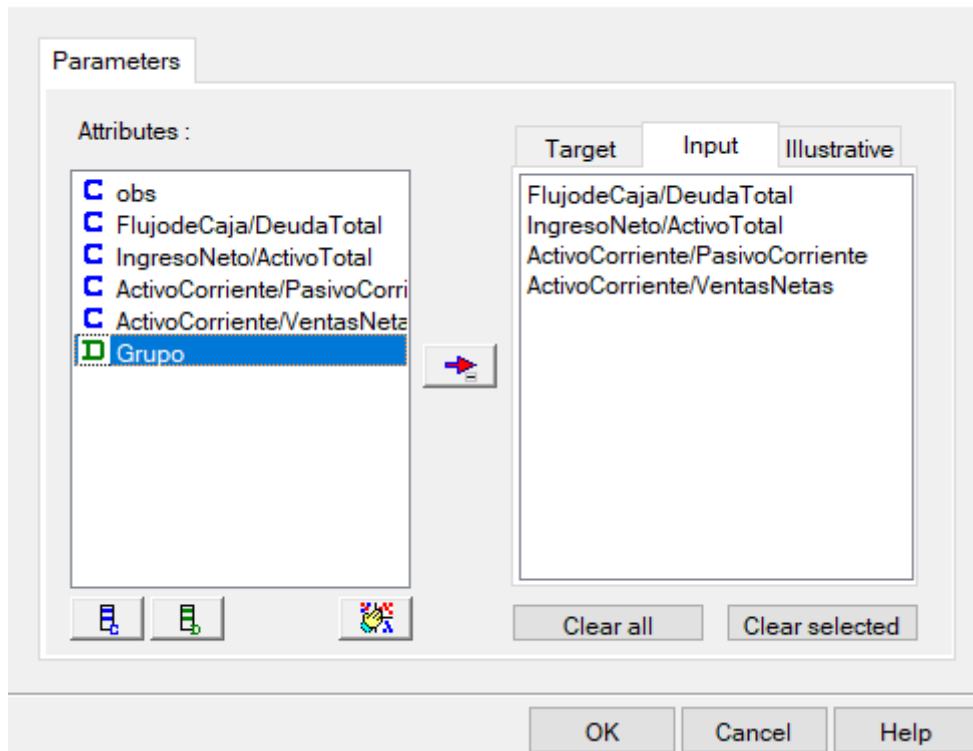
1 y 2.

Vamos a comenzar abriendo la herramienta Tanagra. Vamos a “File→New File y buscamos el archivo empresas.txt, el cual será estudiado.



Vamos ahora a “Define Status” y añadimos como “inputs” todos los atributos salvo “obs” y “Grupo”. El atributo “Grupo” lo usaremos como “target”.

Define attribute statuses



Una vez hecho esto, pulsamos “Ok” y ejecutamos nuestro “Define Status”, asegurándonos de que los atributos quedaron clasificados correctamente.

Default title
Dataset (empresas.txt)
Define status 1

Define status 1

Parameters

Target : 1
Input : 4
Illustrative : 0

Attribute	Target	Input	Illustrative
obs	-	-	-
FlujodeCaja/DeudaTotal	-	yes	-
IngresoNeto/ActivoTotal	-	yes	-
ActivoCorriente/PasivoCorriente	-	yes	-
ActivoCorriente/VentanasNetas	-	yes	-
Grupo	yes	-	-

Computation time : 0 ms.
Created at 20/05/2020 1:20:26

3.

En la pestaña “Spv learning”, que podemos encontrar en la parte inferior de la pantalla, seleccionaremos la herramienta de aprendizaje supervisado “SVM” y la arrastraremos hasta nuestro “Define Status”.

Default title
Dataset (empresas.txt)
Define status 1
Supervised Learning 1 (SVM)

Supervised Learning 1 (SVM)

Parameters

SVM Parameters	
Exponent	1
Filter type	NORMALIZE
Use polynom space normalization	0
Use RBF kernel	0
Gamma for RBF kernel	0,0100
Complexity	1,0000
Calculation parameter	
Epsilon for rounding	1,0E-012
Tolerance for accuracy	1,0E-003

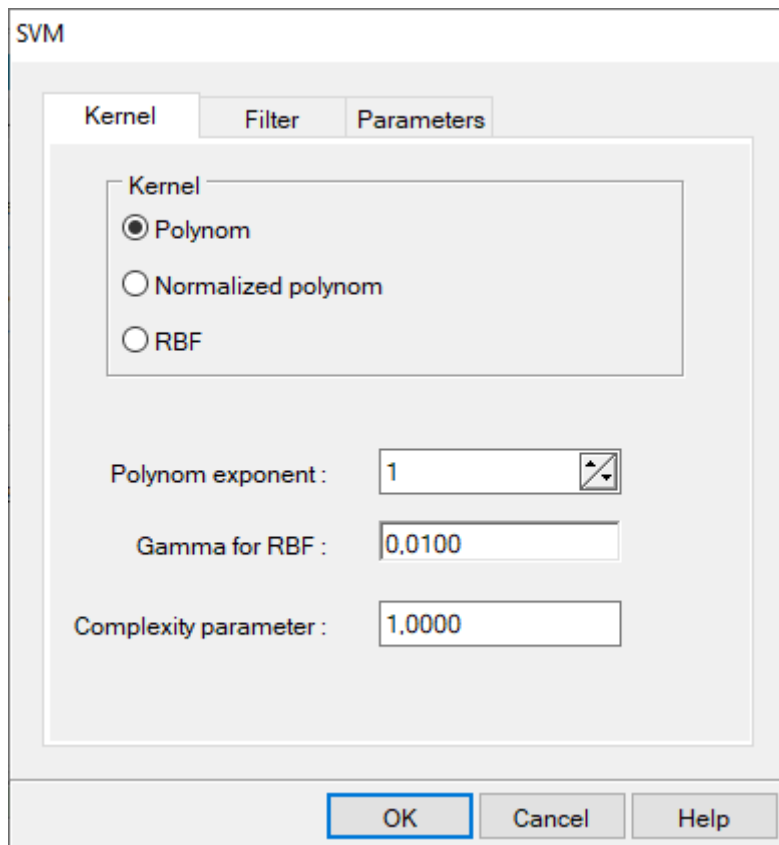
Results

Classifier performances

Error rate		0,1739	
Values prediction		Confusion matrix	
Value	Recall	1-Precision	
Quebró	0,6667	0,0667	
No quebró	0,9600	0,2258	

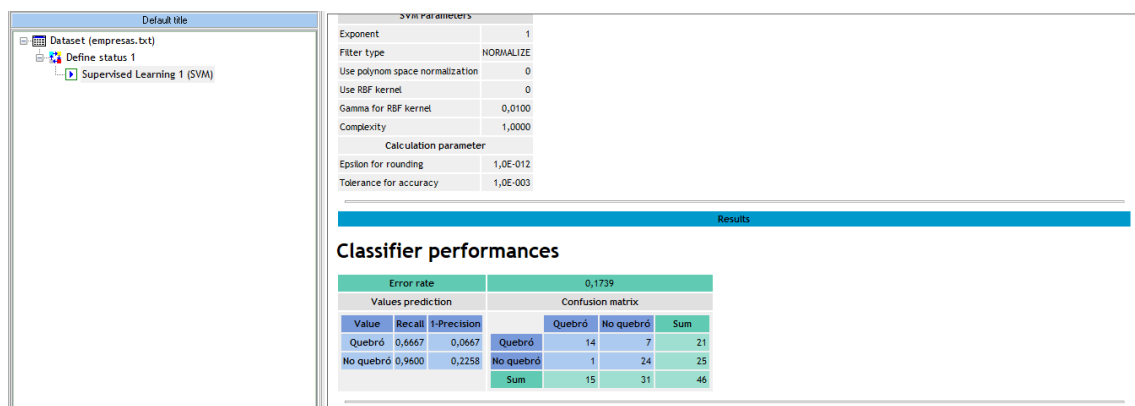
	Quebró	No quebró	Sum
Quebró	14	7	21
No quebró	1	24	25

Si pulsamos con el botón derecho en “Supervised Learning 1 (SVM)” y seleccionamos la opción “Supervised parameters..” llegaremos a una ventana en la que podemos observar que en el exponente del polinomio es 1. Esto quiere decir que todos los exponentes de las variables de entrada serán 1.



4.

Una vez ejecutamos el algoritmo “SVM”, vemos la siguiente pantalla en la que se recogen los datos obtenido por el algoritmo.



Aquí podemos ver la matriz de confusión, en la que 38 elementos quedaron bien clasificados, es decir, el modelo y la realidad coincidieron. En el caso de que la empresa quebró, 14 elementos quedaron bien clasificados ya que tanto el modelo como la realidad llegaron a la conclusión de que la empresa quebró, y 7 mal clasificados, ya que en realidad no quebró, pero el modelo previó que si quebró.

Lo mismo aplicamos a no quebró, salvo que en esta ocasión solo una empresa fue mal clasificada.

5.

En esta misma pantalla, más debajo de la matriz de confusión, podemos observar los pesos que tienen los atributos, así como el término independiente.

El peso es el coeficiente que multiplica al atributo en la ecuación de una combinación lineal de las variables de entrada, y el término independiente es la constante en dicho polinomio:

The screenshot shows the Orange3 interface. On the left, a workflow is visible with 'Dataset (empresas.txt)', 'Define status 1', and 'Supervised Learning 1 (SVM)'. The main panel displays 'Classifier characteristics' for the SVM model. It includes a 'Data description' section showing the target attribute as 'Grupo (2 values)' and 4 descriptors. Below this, the 'Linear classifier' section shows the 'Reference' class value as 'No quebró'. A table lists the weights for each attribute, and the computation time and creation date are shown at the bottom.

Attribute	Weight
FlujodeCaja/DeudaTotal	1,329988
IngresoNeto/ActivoTotal	2,809264
ActivoCorriente/PasivoCorriente	0,449724
ActivoCorriente/VentasNetas	-0,610591
constant	-0,338902

Computation time : 0 ms.
Created at 20/05/2020 1:28:23

Por lo tanto el término independiente será -0,338902

6.

Por último, en la pestaña “Data visualization”, seleccionamos “View dataset” y lo arrastramos y colgamos debajo de “Supervised Learning 1 (SVM)”. Una vez hecho esto lo ejecutamos llegando a la siguiente pantalla:

The screenshot shows the Orange3 interface with the 'View dataset 1' widget added to the workflow. The widget displays a table with 24 rows of data, including observed and predicted values for the 'Grupo' attribute.

	obs	FlujodeCaj	IngresoNet	ActivoCorr	ActivoCorr	Grupo	pred_SpvIn
1	8	-0,0653	-0,0566	1,3737	0,4032	Quebró	Quebró
2	29	-0,0173	0,0233	2,0538	0,3484	No quebró	No quebró
3	35	0,1933	0,0473	2,2506	0,3309	No quebró	No quebró
4	40	-0,333	-0,0854	3,0124	0,473	No quebró	No quebró
5	42	0,5603	0,1112	4,2918	0,4443	No quebró	No quebró
6	24	0,3776	0,1075	3,2651	0,3548	No quebró	No quebró
7	41	0,4785	0,091	1,2444	0,1847	No quebró	No quebró
8	13	0,0109	0,0011	2,1495	0,6469	Quebró	No quebró
9	11	-0,2298	-0,2961	0,331	0,1824	Quebró	Quebró
10	20	0,1227	0,1055	1,1434	0,1655	Quebró	No quebró
11	38	0,2907	0,0597	1,8381	0,3786	No quebró	No quebró
12	34	0,1398	-0,0312	0,4611	0,2643	No quebró	Quebró
13	23	0,0769	0,0195	2,0069	0,5304	No quebró	No quebró
14	19	0,0115	-0,0032	1,2602	0,6038	Quebró	Quebró
15	22	0,5135	0,1001	2,4871	0,5368	No quebró	No quebró
16	45	0,1661	0,0351	2,4527	0,137	No quebró	No quebró
17	27	0,3132	0,0511	4,45	0,6852	No quebró	No quebró
18	3	0,0643	0,0156	1,0077	0,3978	Quebró	No quebró
19	10	-0,1353	-0,1433	1,4196	0,4347	Quebró	Quebró
20	35	0,1379	0,0728	2,6123	0,5151	No quebró	No quebró
21	16	0,3703	0,1098	1,9941	0,3828	Quebró	No quebró
22	9	0,0724	-0,0076	1,3723	0,3361	Quebró	No quebró
23	21	-0,2843	-0,2703	1,2722	0,5128	Quebró	Quebró
24	33	-0,0985	-0,0123	2,5029	0,5778	No quebró	No quebró

Observamos que las dos ultimas columnas son grupo y la previsión de grupo respectivamente. Aquí podemos comprobar, en la última columna, si el algoritmo predijo correctamente si una empresa quebró o no quebró.