

ANNEX I. Xarxes neurals artificials (*Artificial Neural Networks, ANN*)

Les xarxes neurals (*neural networks*) són un tipus d'eina computacional que funciona de forma anàloga als processos biològics que es donen al cervell. Comprenen un conjunt d'unitats de còmput simples però altament interconnectades que s'anomenen nodes, i les qual són responsables de dur a terme uns pocs i rudimentaris còmputs. Els nodes s'organitzen en una sèrie de capes interconnectades a través d'unions funcionals (Figura 1).

Una de les principals característiques de les xarxes neurals és la seva capacitat d'aprendre sistemes complexes. És a dir, si se li dóna un conjunt seqüencial d'*inputs* (variables classificatòries) i un *output* conegut per a un sistema determinat, la xarxa s'organitza internament de tal manera que pot predir de forma acurada l'*output* esperat

per a qualsevol altre *input*. El procés intern d'auto-organització o desenvolupament d'una representació general del sistema es coneix com a entrenament, el qual és crucial per a una fase predictiva eficient de la xarxa neural. Es considera que una xarxa neural està ben entrenada si la desviació entre l'*output* predit per aquesta i l'*output* real es troba dins un límit tolerable. Un dels factors que afecta al procés d'entrenament és el nombre de nodes empleats en la capa oculta. Si n'hi ha pocs, pot ser que no es donin prou oportunitats a la xarxa neural per capturar les intricades relacions entre *inputs* i *outputs*, com per exemple els paràmetres indicadors i l'absència o presència de norovirus en mol·luscs bivalves (Brion i col., 2003). D'altra banda, massa nodes no només suposen un temps de còmput molt llarg per dur a terme un entrenament acurat, sinó que

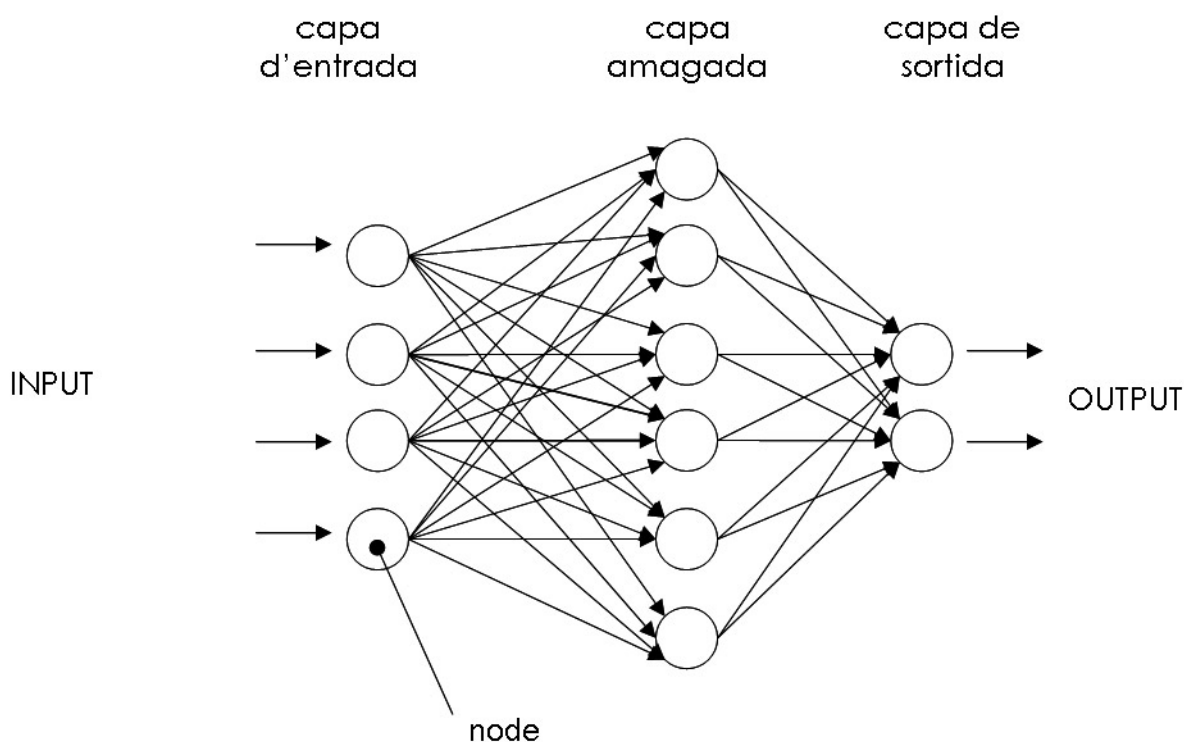


Figura I. Esquema d'una xarxa neural típica (adaptat de Brion i Lingireddy, 1999)

poden provocar un "sobre-entrenament". El sobre-entrenament ocorre quan la xarxa es centra més en les característiques individuals de cada punt de dades que en els patrons generals presents en el conjunt sencer que s'utilitza en l'entrenament.

Generalment, les xarxes neurals *feed-forward* s'entrenen amb un algoritme de *back-propagation*. Aquest algoritme utilitza un mètode de gradient descendent per a minimitzar errors. Començant amb un conjunt assumit o especificat de pesos per a cadascuna de les connexions, l'algoritme de *back-propagation* calcula l'error entre els valors d'*output* predits i els reals especificats en la capa de sortida. Si aquest error és superior a una certa tolerància, l'algoritme actualitza els pesos de les connexions portant l'error cap enrera mitjançant el mètode de gradient descendent. La informació que es transmet d'una capa a la següent experimenta un transformació basada en la següent funció sigmoïdal de transferència:

$$O = \frac{1}{1 + \exp(-i)}$$

on i és l'*input* donat a un node de la xarxa neural i O és l'*output* corresponent a aquell node. De l'equació es desprèn que l'*output* sempre es trobarà entre 0 i 1. Per ensamblatge d'una base de dades amb valors coneguts d'*inputs* i *outputs*, la xarxa es pot entrenar i després utilitzar per a classificar nous conjunts de dades. Referint-nos a l'exemple anterior, a l'*output* absència de norovirus se li assignaria un 0 i al de presència un 1.

Aplicació de les xarxes neurals en l'anàlisi dels resultats obtinguts en el present estudi

Brion i col. (2003) han analitzat la matriu de dades generada en aquesta tesi pel sistema de xarxes neurals artificials per tal de determinar si és possible predir la presència o absència de norovirus en mol·luscs bivalves. També es seleccionaren les dades

corresponents a Suècia i Espanya per determinar si la importància relativa de les variables (*input*) és similar.

El model ensinistrat de xarxa neural identifica el 73,9% de mostres positives per NV. En aquest model de xarxa neural, la classificació absència de NV és del 97,2% i per la taxa de classificació correcta general és 93,4%. A més, troba que la variable País és significativa, fet d'esperar degut a les diferències entre els nivells d'indicadors i certs fenòmens climàtics ocorreguts durant el mostreig.

En el model de xarxa neural, la variable Mes té la major RSE (Relative Strength Effect), valor que indica la importància relativa de les variables o *inputs* per a predir els *outputs* (presència o absència de norovirus). Cal dir que la variable Mes pot relacionar-se amb la temperatura de l'aigua de mar, la supervivència de patògens i indicadors i fenòmens climàtics. En el cas de la xarxa neural, la inclusió de la variable Temperatura millora la capacitat predictiva, mentre que en el model de regressió logística aplicat en el capítol II la disminueix. És important destacar que ambdues anàlisis coincideixen en què tant Fags F-ARN i Colifags somàtics són les variables de major importància a l'hora de predir la presència o absència de norovirus.

Per a investigar les diferències específiques de país, Brion i col. apliquen el model de xarxes neurals separatament a les dades obtingudes a Espanya i Suècia. Aquest paísos són els seleccionats ja que mostren els models de xarxa neural més forts (>80% de predicció de presència de NV i >90% d'absència), tot i ser molt diferents en quant a la mitjana de concentracions d'indicadors i a fenòmens climàtics. Al comparar els respectius RSE es veu que, per Suècia, les variables de major importància són Àrea, Fags de *Bacteroides fragilis* i Mes, mentre que *E. coli* té poca importància relativa. En el model d'Espanya, en canvi, les variables de major importància són Colifags somàtics i Fags F-ARN, seguits per Mes i *E. coli*. Així doncs, veiem que la relació subjacent entre indicadors i norovirus és diferent quan s'analitzen les nostres dades per xarxes neurals.

ANNEX II. Altres publicacions relacionades amb aquesta tesi



ARTICLES

Bofill-Mas S., **M. Formiga-Cruz**, P. Clemente-Casares, F. Calafell, R. Girones. Potential transmission of human polyomaviruses through the gastrointestinal tract after exposure to virions or viral DNA. 2001. *Journal of Virology*; 75(21): 10290-10299.



COMUNICACIONS ORALS

Bofill-Mas, S., N. Albiñana-Giménez, **M. Formiga-Cruz**, R. Girones. Potential transmission of human polyomaviruses through water or food contaminated with urban sewage. *Symposium on Health-Related Water Microbiology*, 14-19 Setembre 2003.

Bofill-Mas, S., P. Clemente-Casares, A. Hundesa, **M. Formiga-Cruz**, R. Girones. Virus emergentes en agua y en moluscos bivalvos. *XIII Congreso de Microbiología de los Alimentos*, Bilbao 16-19 setembre 2002.

Formiga-Cruz, M., Lees, D.N.; Henshilwood, K., Allard, A., Hernroth, B., Vantarakis, A., Papapetropoulou, M., and Girones, R. European study of viral pollution in shellfish: potential indicators. *4th International Conference of Molluscan Shellfish Safety*. Santiago de Compostela, Juny de 2002.



PÒSTERS

Formiga-Cruz M., D.N. Lees, K. Henshilwood, A.K. Allard, B.E. Hernroth, A. Vantarakis, M. Papapetropoulou, R. Girones. Contaminación vírica de moluscos bivalvos en europa: posibles indicadores. *XIII Congreso de Microbiología de los Alimentos*, Bilbao 16-19 setembre 2002.

Formiga-Cruz, M., C. Lleti, M. D. Furones, R. Girones. Relación entre la actual clasificación de las áreas de producción de moluscos bivalvos y la contaminación viral. *XIII Congreso de Microbiología de los Alimentos*, Bilbao 16-19 setembre 2002.