

Objetivo: Utilizar redes neuronales para predecir la supervivencia de pacientes afectados con cáncer de mama y comparar sus resultados con otros métodos tradicionales.

Conjunto de datos: Datos de microarray de ADN utilizados en el estudio de Van't Veer et al. [P2R1][P2R3].

Bibliografía:

[P2R1] Van 't Veer L.J., Dai H., Van de Vijver M.J., et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415, pp. 530-536 (2002).

[P2R2] Van de Vijver M.J. , He Y.D., Van't Veer L.J., et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*, 347, pp. 1999-2009 (2002).

[P2R3] The seventy genes data R package.

[P2R4] Jerez, J.M., Franco, L., Alba, E., et al. Improvement of Breast Cancer Relapse Prediction in High Risk Intervals Using Artificial Neural Networks. *Breast Cancer Research and Treatment*, 94, pp. 265--272 (2005).

Detalles:

a) Del conjunto original seleccionar utilizando algún método sencillo alrededor de 70 genes para utilizarlos en predicción. Comparar la selección con el conjunto de 70 genes seleccionado en la Ref. [P2R1].

b) Diseñar arquitecturas de redes neuronales para predecir la supervivencia utilizando los dos conjuntos de genes mencionados en a). Utilizar una división de patrones similar a la utilizada en [P2R1], para estimar la bondad del método, utilizando un conjunto de validación para ajustar los parámetros del algoritmo, de la arquitectura y para evitar el sobre-entrenamiento. Analizar las peculiaridades del problema de predicción de supervivencia y los modelos existentes [P2R4].

c) Analizar y comparar los resultados obtenidos utilizando un modelo alternativo (regresión logística, SVM, perceptrón simple) y con resultados publicados, analizando la capacidad de generalización, curva ROC, desviación estándar, predicción a distintos tiempos, etc.

d) Generar un informe sobre las tareas realizadas resultados obtenidos, adjuntando el código utilizado.