

Practica 4

Detección de grupos de genes en datos de Microarrays

A partir de los datos proporcionados en la práctica 3, sobre datos de Microarrays de pacientes con cáncer, y después del proceso de sustitución de datos perdidos mediante el método de los ($k = 15$) vecinos más cercanos (proceso ya aplicado en la práctica 3), se desea realizar la eliminación de datos planos del fichero Datos_Microarrays_Cancer.txt. y aplicar una red de Kohonen para obtener grupos de genes coexpresados.

Eliminación de datos planos

A veces aparecen datos flat, datos que son constantes para todos los experimentos (cercaos al cero), estos datos pueden producir ruido y dificultar el procesamiento posterior, en el sentido de que cuando se realiza un proceso de cluster, si utilizamos medidas de correlación, dichos datos pueden ser asignados erróneamente a cualquiera que esté relativamente cercano.

Una de las formas de resolver este problema, es la eliminación de elementos con menos de un número prefijado de “picos” (valores por encima o por debajo de un umbral, en definitiva máximos o mínimos). Los parámetros los suele escoger el usuario. Así (0.2, 0.4, 0, 0.1, -0.3, 4, 0.6, -7, 0.8) tendría dos “picos” si se considera que el umbral es 1

Actividades:

- 1.- Implementar un programa que devuelva el fichero **resultado.txt**, obtenido al aplicar la eliminación de datos planos de cáncer DLBCL del fichero Datos_Microarrays_Cancer.txt. Tomar como máximo número de picos, 7 con un umbral 1.0, es decir eliminar los genes que tengan como máximo (0 ó 1 ó 2... ó 7) variables, cuyo valor esté fuera del intervalo $[-1, +1]$
- 2.- Detectar los grupos que aparecen de genes, aplicando una red de Kohonen.
- 3.- **(Opcional)** Después de obtener el mapa topológico mediante la aplicación de la red de Kohonen a los datos procesados, se deberá estudiar la forma de obtener los cluster, aportando la documentación que justifique dicho proceso, y su aplicación al caso concreto del mapa topológico obtenido.

Nota: La implementación de la red Kohonen que se puede utilizar no estará restringida a MatLab, sino que puede ser utilizado cualquier otro software libre como las SomToolbox del propio Kohonen, o la SOM del servidor GEPAS (<http://gepas.bioinfo.cnio.es/>) .

.

Entregar:

Software

- 1.- Los programas fuente correspondiente para la obtención del ejecutable de eliminación de datos planos.
- 2.- El fichero de texto resultado.txt obtenido en la actividad 1.

Documentación en formato Word, Postscript o PDF

- 1.- Estudio teórico detallado de la red de Kohonen, que demuestre que se ha entendido su funcionamiento (**El estudio teórico, es muy importante para que la práctica sea superada**).
- 2.- Resultados de la aplicación de la Red de Kohonen al problema considerado.
- 3.- **(Opcional)** Justificación detallada, así como las referencias de los documentos en los que se basa la obtención de los cluster obtenidos (número de cluster obtenido, número de elementos de cada cluster, y elementos (números) que pertenecen a cada uno de los cluster).

La Fecha de entrega de la práctica, será conjuntamente con la práctica 3, hasta el 6 de Junio de 2014