

Seminario de seguimiento 2

Seminario de seguimiento 2	
Asistentes	Patricio García Báez, Oscar Trujillo Acosta y Alberto Fariña Barrera
Fechas	13 de abril de 2016 a 25 de mayo de 2016
Hora	12:30
Temas tratados	Dificultades encontradas y pasos a seguir

1 INTRODUCCIÓN

En este fichero se tratarán las reuniones que tuvieron lugar entre las fechas indicadas anteriormente en el despacho del profesor Patricio García una vez cada dos semanas. En ellas se trata cualquier problema que se tenga y se marcan hitos para continuar con el desarrollo de manera ordenada.

2 PASOS A SEGUIR

Ya obtenidas las primeras pruebas de la red neuronal para cada tipo de cáncer, en esta etapa comienza la fase de análisis, en la cual es necesario realizar pruebas para distintos valores de las variables importantes para decidir con cuál de ellos se realizará la red neuronal final.

Para el Back-Propagation, que es el algoritmo con el que se han realizado las primeras pruebas se tendrán que modificar 3 variables clave:

- Nº de neuronas ocultas
- Learning Rate
- Momentum

Posteriormente se realizarán pruebas para otros modelos de red neuronal, tales como Supported Vector Machine o aprendizaje no supervisado con Deep Belief Networks, con el objetivo de comprobar con qué algoritmo se obtienen mejores resultados.

3 DIFICULTADES ENCONTRADAS

Se han encontrado varias dificultades durante esta fase:

- Es necesario hacer un gran número de pruebas, lo que conlleva bastante tiempo de análisis, esto se ha solucionado elaborando scripts que realicen todas las pruebas de una vez, exportando la salida a ficheros 'csv' para su posterior importación a ficheros Excel para elaborar gráficas y tablas, reduciendo sobremedida el tiempo necesario para la realización de las pruebas.

- La librería PyBrain no muestra por defecto los porcentajes de aciertos al validar la red neuronal, por lo que fue necesario elaborar de nuevo un script que obtuviera el valor de creencia para las dos salidas y lo asignara a las categorías 'positivo', 'negativo', 'falso positivo' y 'falso negativo', para posteriormente obtener los porcentajes y los valores de las variables como accuracy o recall.