

UNIVERSIDAD DE GRANADA

PROYECTO FINAL:

APRENDIZAJE AUTOMÁTICO

Parkinson Telemonitoring

Autor:

Luis Suárez Lloréns

June 23, 2016

Contents

1	Motivación	2
2	Información de los datos	2
3	Selección de modelos	3
4	Manipulación de los datos	4
4.1	Codificación de los datos	4
4.2	Normalización	4
4.3	Selección de características	4
5	Resultados	4

1 Motivación

La medicina se ha visto impulsada en los últimos años por los avances técnicos. En concreto, se han utilizado técnicas de inteligencia artificial para la ayuda en el diagnóstico y seguimiento de enfermedades.

El objetivo no es sustituir la función de los profesionales del campo de la medicina, si no crear herramientas para mejorar el tratamiento a los pacientes.

Este trabajo se centra en el seguimiento de la enfermedad de Parkinson, y en obtener una estimación del grado en la que esta afecta al paciente de manera automática, mediante la medición de la voz del mismo. Además, estas mediciones se pueden realizar de forma automática en la casa del paciente.

Todo esto nos permitiría realizar un mejor seguimiento del paciente, y reduciendo posibles inconvenientes, como los desplazamientos por parte del paciente a su centro médico y la realización de pruebas médicas más costosas, tanto económicamente como en tiempo de personal.

2 Información de los datos

La base de datos, creada por los investigadores Athanasios Tsanas y Max Little de la universidad de Oxford en colaboración con varios centros médicos de los Estados Unidos, contiene datos relativos al habla de varios enfermos de Parkinson.

En la creación de la base de datos, 42 pacientes fueron monitorizados en un periodo de 6 meses. En total, se realizaron 5,875 mediciones a los pacientes.

En la investigación se trata de conocer el grado de Parkinson del paciente, utilizando para ello la escala unificada para la evaluación de la enfermedad de Parkinson (UPDRS). Por tanto, el objetivo es utilizar los 16 valores obtenidos de la medición del habla para estimar los valores de las variables "motor UPDRS" y "total UPDRS".

En la tabla de los datos también se puede consultar el identificador del paciente al que se realizó la medición, su edad, género y el momento exacto de su realización, pero el estudio original ignora estos factores.

3 Selección de modelos

Ya hemos visto que el objetivo de este estudio es estimar dos valores reales. Nos centraremos para la selección de modelos en los métodos de regresión que hemos estudiado durante la asignatura.

Utilizaremos como una primera aproximación la regresión logística. Esta técnica, básica en el desarrollo del aprendizaje automático, sigue siendo de validez y puede obtener buenos resultados. Además, dada la simplicidad de la técnica, podemos ajustar varios modelos —lineales, cuadráticos, ...— y utilizar estos resultados como medidas para comparar los resultados obtenidos con técnicas más complejas. Destacar el posible inconveniente de tener que elegir las familias de funciones manualmente.

Otra aproximación posible es usar un modelo de función de base radial. Parece lógico pensar que dos mediciones con valores parecidos nos darán valores de UPDRS parecidos. Para poder usar esta técnica, además de la selección del núcleo, tenemos que preocuparnos de manipular los datos, normalizándolos, pues los resultados de las mediciones no se encuentran todos en la misma escala.

Por último, utilizaremos una modelo de red neuronal multicapa. La potencia y principal bondad de este tipo de modelos es su capacidad de aprender automáticamente posibles relaciones ocultas de gran complejidad de oculta detrás de los datos. Podríamos entender esto como una especie de autoselección de características. Esto es importante en el ámbito de la medicina, donde los parámetros están interconectados, posiblemente mediante procesos muy complejos de muy difícil modelización en ecuaciones. El problema de usar este tipo de técnicas es la selección de la arquitectura —número de neuronas, distribución de las neuronas en capas, tipo de funciones no lineales, ...—.

4 Manipulación de los datos

4.1 Codificación de los datos

Los datos obtenidos directamente se encuentran en una codificación perfecta para empezar a trabajar. Se tratan de 18 columnas —una vez eliminadas las columnas de identificador, edad,...— de valores reales, dos de ellas son los valores a estimar, y el resto las mediciones de la voz de los pacientes.

4.2 Normalización

4.3 Selección de características

5 Resultados