Trabajo Práctico

2do cuatrimestre 2024 — Metodos Numericos

Integrante	LU
Francisco Fazzari	900/21
Felipe Miriuka	1693/21
Ignacio Fernández Oromendia	29/21

Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

Ciudad Universitaria - Pabellón I

Intendente Güiraldes 2160 - C1428EGA

Ciudad Autónoma de Buenos Aires - Argentina

Tel/Fax: (54 11) 4576-3359

http://exactas.uba.ar

Índice

1. Introducción Teórica

En este informe evaluaremos el algoritmo de cuadrados mínimos sobre datos de series temporales de Electrocardiogramas. Vamos a evaluar los datos usando el algoritmo clásico de cuadrados mínimos y el algoritmo de cuadrados mínimos con regularización. Para los experimentos evaluaremos como mejora o empeora el error en base al grado de los polinomios usados en el algoritmo de cuadrados mínimos.

2. Desarrollo

2.1. Cuadrados mínimos

El problema de cuadrados mínimos [1] se define como $\|X\beta - y\|_2^2$ se utiliza para aproximar una función a una muestra de datos. Podemos ver la norma como $\beta = (X^tX)^{-1}X^ty$ y utilizando la descomposición SVD la ecuación nos queda de la siguiente manera, $\beta = V\Sigma^{-1}U^ty$. Veamos como se vería el algoritmo del siguiente problema.

Algorithm 1 Cuadrados Mínimos (X, y)

$U\Sigma V^t = SVD(X)$
$\Sigma^{-1} = invertirDiagonal(\Sigma)$
return $V\Sigma^{-1}U^ty$

Podemos aprovechar que Σ es una diagonal y por lo tanto para calcular su inversa se podemos asignar en $(\Sigma^{-1})_{ii} = 1/(\Sigma)_{ii}$.

2.2. Cuadrados mínimos con regularización

La ecuación de cuadrados mínimos sin regularización resuelve el problema pero puede suceder que los valores de β sean muy grandes, lo cual puede no ser bueno para predecir nuevos puntos (puntos que no se encontraban en la muestra). Para ello vamos a usar cuadrados mínimos con regularización, donde se modifica la ecuación original para también minimizar los β.

Ahora los β se formulan de la siguiente manera:

$$\beta = (X^tX + \lambda I)^{-1}X^ty$$

Si tomamos la descomposición SVD de X como $U\Sigma V^t$ podemos desarrollar la fórmula de β:

$$\begin{aligned}
\beta &= (X^tX + \lambda I)^{-1}X^ty \\
\beta &= ((U\Sigma V^t)^tU\Sigma V^t + \lambda I)^{-1}(U\Sigma V^t)^ty \\
\beta &= (V\Sigma^tU^tU\Sigma V^t + \lambda I)^{-1}V\Sigma^tU^ty \\
\beta &= (V\Sigma^t\Sigma V^t + \lambda VV^t)^{-1}V\Sigma^tU^ty \\
\beta &= (V(\Sigma^t\Sigma + \lambda I)V^t)^{-1}V\Sigma^tU^ty \\
\beta &= (V^t)^{-1}(\Sigma^t\Sigma + \lambda I)^{-1}V^{-1}V\Sigma^tU^ty \\
\beta &= V(\Sigma^t\Sigma + \lambda I)^{-1}\Sigma^tU^ty \\
\beta &= V(\Sigma^2 + \lambda I)^{-1}\Sigma^tU^ty \\
\beta &= V\Sigma(\Sigma^2 + \lambda I)^{-1}U^ty
\end{aligned}$$

Recordamos que tanto V como U son ortogonales, por lo tanto, $VV^t = I$ y $V^{-1} = V^t$, igual para U. Además, como Σ es diagonal, $\Sigma^t = \Sigma$.

Ahora que desarrollamos la forma matricial del problema veamos el algoritmo para resolverlo.

Algorithm 2 Cuadrados Mínimos (X, y)

$U\Sigma V^t = SVD(X)$
$inv = invertirDiagonal(\Sigma^2 + \lambda I)$
return $V\Sigma \times inv \times U^t y$

3. Resultados

3.1. Predicción con polinomio de Legendre

En este experimento compararemos el ECM (error cuadrático medio) de los datos de ajuste con los de validación de un sujeto, y veremos cómo cambia este en función del grado del polinomio usado en el algoritmo de cuadrados mínimos.

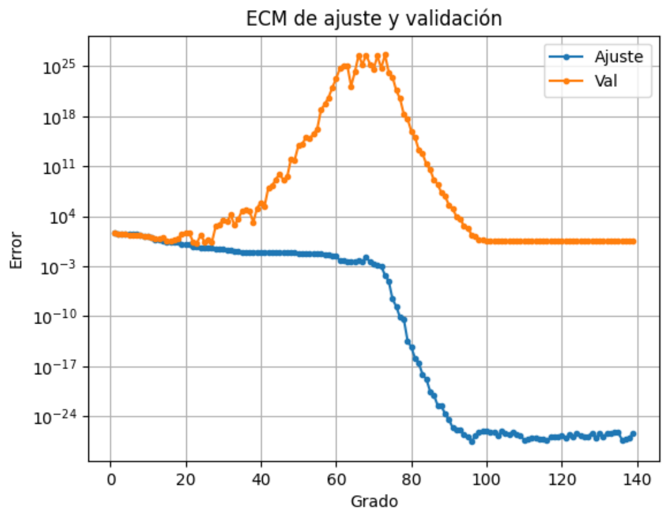

Figura 1: Error en función del grado del polinomio para un sujeto

Se puede observar en el gráfico como al principio el error de los dos sets de datos es muy similar, pero a partir del grado 20 del polinomio el error de los datos de ajuste va bajando levemente mientras que el de los datos de validación aumenta considerablemente. Esto se debe a que comienza a ver sobreajuste sobre los datos de ajuste, y entonces el algoritmo pierde generalidad porque está entrenado para ser muy bueno prediciendo los datos con los que fue entrenado.

Pero a partir del grado 78 se puede observar como el error de ajuste empieza a bajar considerablemente, pero también empieza a bajar considerablemente el error de validación, hasta que a partir del grado 100 el error de ajuste se mantiene muy bajo y el de valuación se mantiene constante con un error muy similar con el cual empezó, lo cual va en contra de lo que uno podría suponer que pasaría por el sobreajuste.

Esto ocurre debido a un fenómeno conocido como doble descenso[1], donde un modelo con una baja cantidad de parámetros y un modelo con gran cantidad de parámetros tendrá un error chico pero un modelo con una cantidad de parámetros similar a la cantidad de datos de la muestra tendrá un error más grande. En el gráfico se puede observar como el peor error de los datos de validación se puede observar alrededor de los polinomios de grado 70, que es la cantidad de datos de muestreo que posee la data utilizada en el experimento.

[1] `https://en.wikipedia.org/wiki/Double_descent`

3.2. Exploración de hiper-parámetros

En nuestro último experimento vamos a explorar los mejores hiper-parámetros para nuestros datos. Para ello vamos a probar con todos los grados para el polinomio de Legendre entre 0 y el doble del largo de los datos. Luego por cada grado vamos a tomar 100 valores distintos de λ espaciados logarítmicamente entre 10^{-8} y 1 y predecir con regularización el electrocardiograma de los 5 sujetos. Para ello utilizamos el algoritmo de predicción desarrollado en la sección anterior y además como son predecir los 5 sujetos con cómputos disjuntos, es decir, no depende uno del otro, utilizamos multiprocesamiento para optimizar la predicción. Esperamos que los resultados tengan una correlación con el fenómeno del doble descenso por lo menos por el lado de los grados. Estos fueron los resultados.

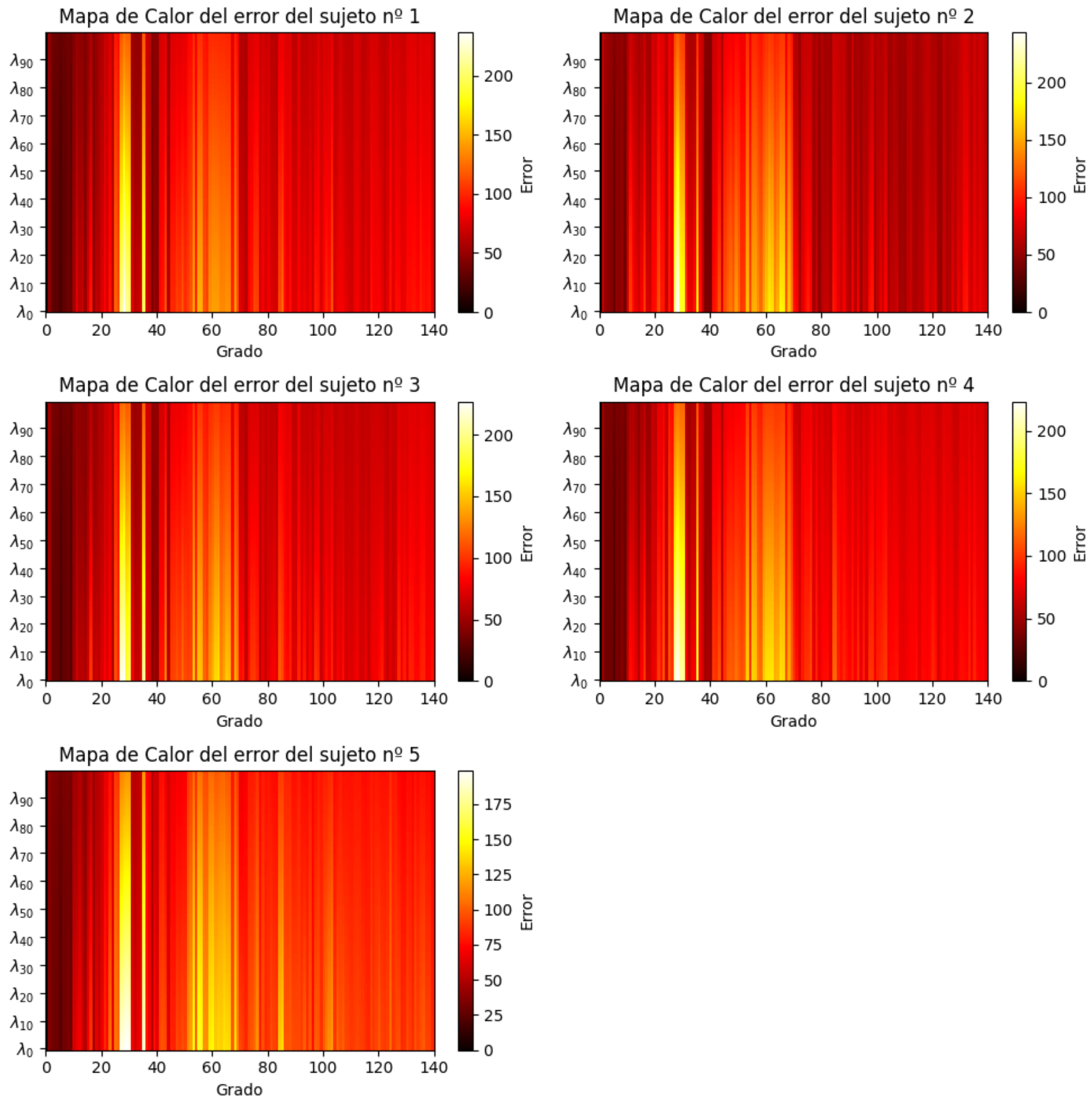

Figura 2: Mapa de calor del error sobre los sujetos

Podemos concluir que para cada sujeto con más o menos intensidad podemos ver el fenómeno del doble descenso donde cerca del grado 30 y para casi todos los valores de λ menos los más cercanos al 1 tenemos un máximo de error el cual después converge a un mismo error para todo λ mayor al error obtenido en los primeros 20 grados. Por otro lado, pudimos estudiar los hiper-parámetros óptimos para cada sujeto. Para nuestra sorpresa, fue el mismo valor para todos, un polinomio de grado 5 y $\lambda = 1$.

4. Conclusiones

En conclusión pudimos experimentar sobre el problema de aproximación con cuadrados mínimos utilizando dos implementaciones distintas que resuelven el problema. Pudimos experimentar con datos reales sobre electrocardiogramas y pudimos predecir con bastante fidelidad el de 5 sujetos sanos. Lo que nos llevamos del siguiente trabajo es que es posible predecir con un par de muestras nuevos datos de los cuales no teníamos conocimiento.

Referencias

[1] J. Douglas Faires Richard L. Burden. *Análisis numérico*. International Thomson Editores, 7th edition, 2002. Chapter 8.1.