Métodos Numéricos
1er Cuatrimestre 2024
Trabajo Práctico 3

DEPARTAMENTO DE COMPUTACION
Facultad de Ciencias Exactas y Naturales - UBA

Cuadrados mínimos con regularización con SVD e interpolación de Legendre

Introducción

El objetivo de este trabajo es la implementación y experimentación de cuadrados mínimos utilizando SVD y el estudio en particular del ajuste de curvas mediante polinomios de Legendre.

A continuación describimos ciertos conceptos que se van a utilizar para realizar las consignas.

Cuadrados mínimos

Dado un conjunto de pares ordenados de valores (x_i, y_i) para $i = 1, \cdots, m$, buscamos una función $f(x)$ perteneciente a una familia F tal que "mejor aproxime" a los datos. Problema de cuadrados mínimos general:

$$\min_{f \in F} \sum_{i=1}^{m} (f(x_i) - y_i)^2$$

Podemos obtener el problema de cuadrados mínimos lineales, si restringimos a la familia de funciones a que sea la dada por combinaciones lineales de una base con fuciones linealmente independientes, $F = \{f(x) = \sum_{j=1}^{n} \beta_j \phi_j(x)\}$:

$$\min_{\beta_1, \cdots, \beta_n} \sum_{i=1}^{m} (\sum_{j=1}^{n} \beta_j \phi_j(x_i) - y_i)^2$$

Por ejemplo puede utilizarse una base de funciones polinomiales.

Una vez pasada a forma matricial (ver teórica diapo 9). El problema se formula como:

$$\min_{\beta \in \mathbb{R}^n} ||X\beta - y||_2^2$$

y su solución en los casos de tener mayor cantidad de datos que funciones $(m > n)$ está dada a partir de las ecuaciones normales $X^t X \beta = X^t y$ y es:

$$\beta = (X^t X)^{-1} X^t y$$

Esta expresión puede adaptarse si se conoce la descomposición SVD de $X = USV^t$. Si considereamos el caso donde $m > n$ y X tiene rango completo, U y S pueden reducirse para que S quede cuadrada eliminando las filas nulas y se obtiene:

$$\beta = VS^{-1}U^t y$$

Donde invertir la matriz no es costoso ya que es una matriz diagonal. El error total de la regresión queda dado por:

$$ECM = \|X\beta - y\|_2^2$$

Cuadrados mínimos con regularización

En muchas situaciones de optimización de parámetros, se buscan soluciones donde se agregan otros requerimientos sobre la solución, por ejemplo, en este caso se pueden pedir que los coeficientes β no sean arbitrariamente grandes. Esta idea se conoce como regularización. Especificamente vamos a utilizar regularización L2 (cuadrados) que también se conoce como regresión *ridge*.

Formalmente podemos pedir que lo que queremos minimizar incluya un nuevo término que penalice coeficientes muy grandes, es decir:

$$\min_{\beta_1, \cdots, \beta_n} \sum_{i=1}^{m} (\sum_{j=1}^{n} \beta_j \phi_j(x_i) - y_i)^2 + \lambda \sum_{j=1}^{n} \beta_j^2$$

El parámetro λ indica cuan fuerte es esta penalización.

Para derivar las ecuaciones normales en este caso se puede derivar con respecto a β e igualar a 0. Lo que se obtiene es
$(X^tX + \lambda I)\beta = X^t y$, por lo que la solución en este caso será:

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

Podemos expresear nuevamente esta solución en términos de SVD y obtenemos el siguiente resultado:

$$\beta = VS(S^2 + \lambda I)^{-1}U^t y \tag{2}$$

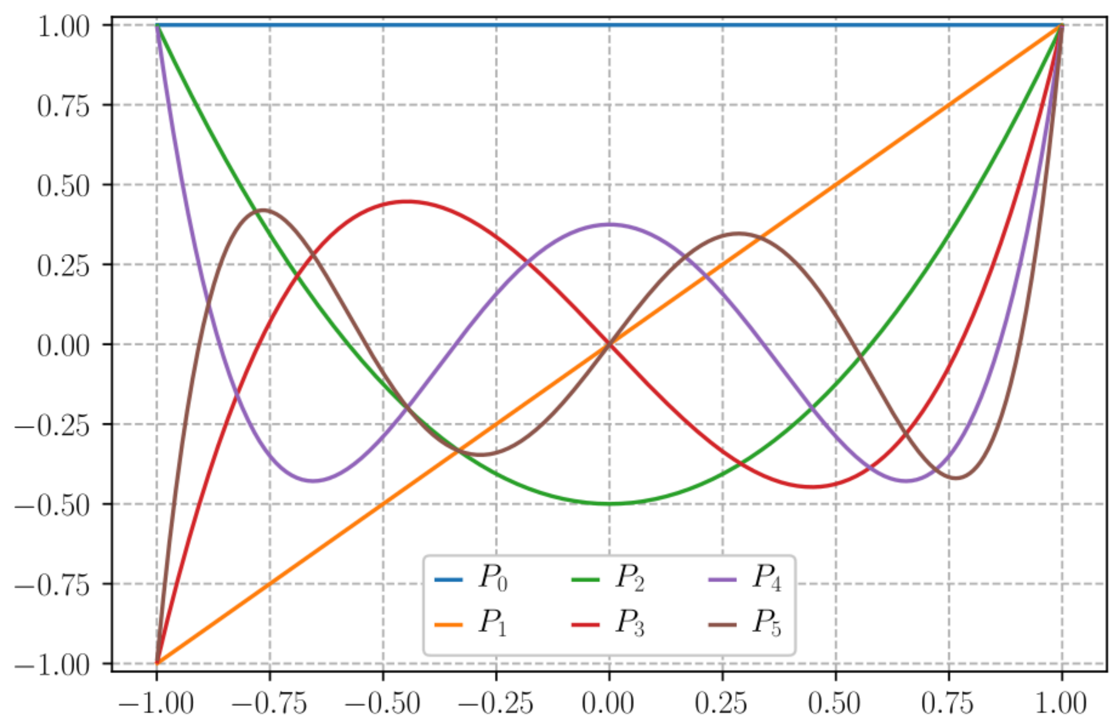
Donde nuevamente lo que hay que invertir es diagonal y puede hacerse trivialmente.

Regresión polinomial con polinomios de Legendre

Se llama regresión polinomial a utilizar una base de funciones linealmente independiente basada, por ejemplo, en las funciones $\{1, x, x^2, \cdots, x^n\}$. Esta base no es la mejor ya que la matriz resultante no tiene buen número de condición [3]. Para mejorar esto se pueden utilizar los polinomios de Legendre que además son una base ortonormal.

Los polinomios de Legendre $P_n(x)$ de grado n tiene la siguiente pinta:

n	$P_n(x)$
0	1
1	x
2	$\frac{1}{2}(3x^2 - 1)$
3	$\frac{1}{2}(5x^3 - 3x)$
4	$\frac{1}{8}(35x^4 - 30x^2 + 3)$
5	$\frac{1}{8}(63x^5 - 70x^3 + 15x)$
6	$\frac{1}{16}(231x^6 - 315x^4 + 105x^2 - 5)$
7	$\frac{1}{16}(429x^7 - 693x^5 + 315x^3 - 35x)$
8	$\frac{1}{128}(6435x^8 - 12012x^6 + 6930x^4 - 1260x^2 + 35)$
9	$\frac{1}{128}(12155x^9 - 25740x^7 + 18018x^5 - 4620x^3 + 315x)$
10	$\frac{1}{256}(46189x^{10} - 109395x^8 + 90090x^6 - 30030x^4 + 3465x^2 - 63)$



Sobreajuste

A partir de un conjuntos de datos X, valores y y la base de funciones, es posible encontrar la mejor solución en términos de cuadrados mínimos, con los mecanismos ya vistos. Una pregunta que se desprende es, que ocurre con esta solución si aparece un nuevo conjunto de datos X_{val} los cuales provienen de la misma distribución de los primeros datos. En general, se puede ver que si se utiliza un grado de polinomio muy alto, o no se utiliza regularización, el error del ajuste aumenta al probar los nuevos datos. A la hora de encontrar el grádo de polinomio y el nivel de regularización adecuados, una buena práctica suele incluir ajustar con ciertos datos y evaluar el error del ajuste en datos no utilizados para ajustar. Este proceso se llama optimización de hiperparámetros con validación cruzada.

El procedemiento de entrenamiento y validación quedaría así:

$$\beta = (X_{ajuste}^t X_{ajuste} + \lambda I)^{-1} X_{ajuste}^t y_{ajuste}$$

$$ECM_{ajuste} = \|X_{ajuste}\beta - y_{ajuste}\|_2^2$$

$$ECM_{val} = \|X_{val}\beta - y_{val}\|_2^2$$

Consignas:

1. A partir de la ecuación 1 derivar la ecuación 2 reemplazando con la definición de SVD.

2. a. Implementar código en python para resolver el problema de cuadrados mínimos utilizando SVD (*np.linalg.svd*).

   b. Implementar código en python para resolver el problema de cuadrados mínimos con regularización utilizando SVD.

3. Utilizando los datos de ajuste y validación provistos, los cuales tienen dos columnas: x e y.

   a. Ajustar una combinación lineal de polinomios de Legendre (sin regularización) en los datos de ajuste y predecir los valores de validación. Calcular los errores de ajuste y de validación para todos los grados desde 1 hasta 2 veces la dimensión de los datos. Graficar ambos errores en función del grado en escala logarítmica en y. Observar el fenómeno de doble desenso (https://en.wikipedia.org/wiki/Double_descent (https://en.wikipedia.org/wiki/Double_descent)).

   b. Explorar los hiperparámetros grado del polinomio y valor de regularización λ. Mostrar con un *heatmap* el error de validación para la grilla de parámetros explorada que capture de forma efectiva la región donde están los mejores parámetros. Reportar el grado y λ que minimiza el error en los datos de validación.

Forma y Fecha de entrega

1. Se debe entregar un informe hecho en Latex. Este debe incluir una introducción al problema, la solución a las consignas mostrando pseudocódigos con una explicación de su funcionamiento. Figuras o tablas para visualizar resultados junto con sus interpretaciones. Las figuras y tablas tienen que estar numeradas y referenciadas en el texto y tener una descripción.
   Una sección de conclusiones con un resumen y de lo mostrado y los principales resultados.

2. Se debe entregar los códigos que generan los resultados, figuras, etc. Si se requieren procedimientos especiales para ejecutar el código deben ser descriptos.

3. La entrega es el **Domingo 30 de Junio 23:58h**. Para la entrega se facilitará un formulario online donde deberan adjuntar el informe y los archivos (incluir en el nombre de los archivos el nombre del grupo).

4. Recuperatorio: **Domingo 14 de Julio hasta las 23.58hs**, enviando el trabajo corregido.

Referencias

[1] Hastie, T., Tibshirani, R., Friedman, J. (2001). The Elements of Statistical Learning. New York, NY, USA: Springer New York Inc.

[2] van Wieringen, W. N. (2015). Lecture notes on ridge regression. arXiv preprint arXiv:1509.09169. pdf (https://arxiv.org/pdf/1509.09169.pdf)

[3] Berztiss, A. T. (1964). Least Squares Fitting of Polynomials to Irregularly Spaced Data. SIAM Review, 6(3), 203–227. http://www.jstor.org/stable/2027588 (http://www.jstor.org/stable/2027588)