

Practica 3 – Regresión Lineal

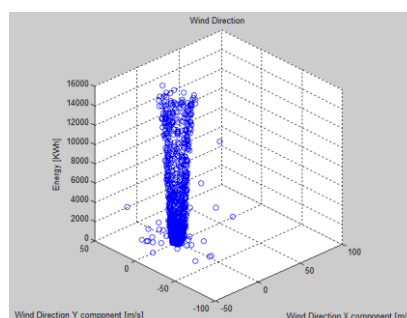
Alumno: Fco. Javier Vargas García-Donas

Resolución

La resolución del ejercicio se ha planteado en el desarrollo de los siguientes ficheros:

P3_Data_Inspection.m

Fichero en el que se muestra la naturaleza de los datos y se muestra que la energía generada no depende de la dirección del viento.



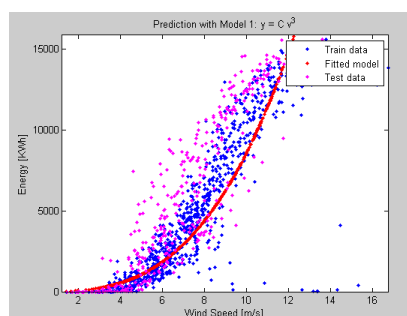
P3_Model_1_previous_knowledge.m

Fichero en el que se modela la energía generada en función del modelo físico que relaciona la velocidad y la energía (por un modelo cúbico).

$$y = C v^3 \quad C = \frac{1}{N} \sum_{i=1}^N \frac{y^{(i)}}{v^3}$$

Con este modelo se obtiene un EV:

EV error of model 1: 0.594812

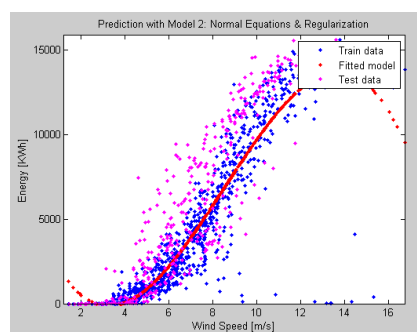


P3_Model_2_1_normal_equations.m

Fichero en el que se modela con una regresión lineal de tercer orden y con regularización.

Con este modelo se obtiene un EV:

EV error of model 2: 0.719952

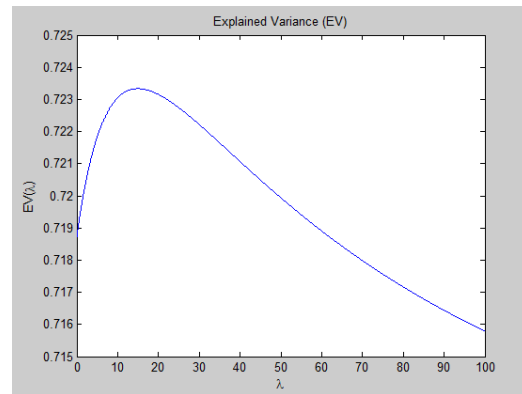


P3_Model_2_2_lambda_error.m

Fichero en el que se varía el valor de la variable de regularización para obtener el mejor valor de EV.

Se aprecia como este valor está en torno a 15. Esto en principio choca un poco ya que el orden del sistema es 3 y no puede hacer mucho sobreajuste así que no haría falta un valor de regularización tan alto.

Se consigue hasta un valor de 0.7235 de EV.



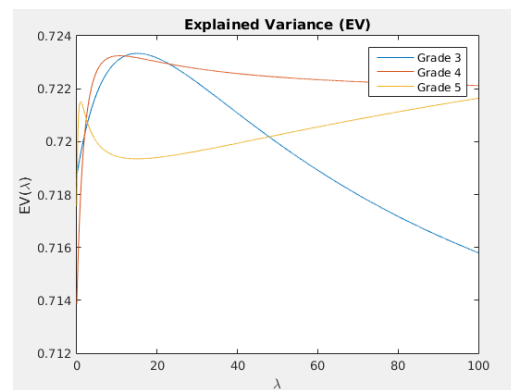
P3_Model_2_3_polynomial_grade.m

Fichero en el que se estudia la validez de diferentes modelos propuestos. Los distintos modelos se diferencian en el orden del polinomio del mismo. Se han generado distintas gráficas donde se observa que el modelo óptimo es el de tercer orden.

P3_Model_2_4_CV.m

Fichero en el que se propone un modelo de validación cruzada. Para ello se juntan los datos de test con los de entrenamiento para luego realizar una nueva separación:

- 20% datos para test
- 80% datos para entrenamiento y validación cruzada de los que distinguimos:
 - 70% para entrenamiento
 - 30% validación cruzada



Dejando los datos de test a parte, se propone un bucle que baraja los datos de entrenamiento y validación cruzada y entrena distintos modelos (para distintas asignaciones de datos de test y validación cruzada). Para cada uno de estos modelos se guardan tanto los pesos como el EV.

Luego se proponen 3 métodos de agregación de los pesos:

1. Hacer una media de los pesos de todos los modelos
2. Usar los pesos del que obtuvo mejor error
3. Usar la media de los pesos de los 3 modelos con menor error

Los resultados varían en cada ejecución, pero casi siempre el enfoque 2 y 3 son mejores que el primero. A pesar de lo que se iba buscando los errores con el modelo agregado son menores que los resultados hallados en el modelo de ecuaciones normales con regularización.

```
EV error of the test data when averaging:    0.189982
EV error of the test data for best EV      :    0.458341
EV error of the test data for 3 best EV    :    0.438745
```