

Prácticas de Aprendizaje Automático

Grupo 3

Trabajo 3: Ajuste de Modelos Lineales

Pablo Mesejo

Universidad de Granada

Departamento de Ciencias de la Computación e Inteligencia Artificial



UNIVERSIDAD
DE GRANADA



Ajuste de Modelos Lineales

- Ajuste y selección del mejor modelo lineal, y estimación del error E_{out} del modelo final
- **Casuística (relativamente) real:** te llega un problema y... ¿cómo lo resuelves?
 - **Análisis** del Problema, **Exploración** de los Datos, Formulación de **Hipótesis**, **Entrenamiento**, **Validación**, y **Discusión** de Resultados

Ajuste de Modelos Lineales

– Problema de clasificación

- <https://archive.ics.uci.edu/ml/datasets/optical+recognition+of+handwritten+digits>

– Problema de regresión

- <https://archive.ics.uci.edu/ml/datasets/airfoil+self-noise>



Optical Recognition of Handwritten Digits Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Two versions of this database available; see folder

Data Set Characteristics:	Multivariate	Number of Instances:	5620	Area:	Computer
Attribute Characteristics:	Integer	Number of Attributes:	64	Date Donated	1998-07-01
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	225870

Source:

E. Alpaydin, C. Kaynak
Department of Computer Engineering
Bogazici University, 80815 Istanbul Turkey
alpaydin@boun.edu.tr



Airfoil Self-Noise Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: NASA data set, obtained from a series of aerodynamic and acoustic tests of two and three-dimensional airfoil blade sections conducted in an anechoic wind tunnel.

Data Set Characteristics:	Multivariate	Number of Instances:	1503	Area:	Physical
Attribute Characteristics:	Real	Number of Attributes:	6	Date Donated	2014-03-04
Associated Tasks:	Regression	Missing Values?	N/A	Number of Web Hits:	96492

Source:

Provide the names, email addresses, institutions, and other contact information of the donors and creators of the data set.
Donor:
Dr Roberto Lopez
robertolopez@intellica.com
Intellica

Creators:
Thomas F. Brooks, D. Stuart Pope and Michael A. Marcolini
NASA

Ajuste de Modelos Lineales

1. **Comprender el problema** a resolver → Visualizar datos!
2. **Preprocesado de los datos** → eliminar datos sin variabilidad, reducción/aumento de dimensionalidad, normalización de datos,...
3. Selección de clases de **funciones a usar** → combinaciones lineales/cuadráticas/etc. de los valores observados
4. Definición de los conjuntos de **training, validación y test** usados en su caso → ¿uso de cross-validation?
5. Discutir la necesidad de **regularización** y, en su caso, la función usada para ello.
6. Definir los **modelos LINEALES a usar y estimar sus parámetros**
7. Selección y **ajuste del modelo final**.
8. Discutir la idoneidad de la **métrica** usada en el ajuste → MSE, MAE, Cross-Entropy,...
9. **Estimación del error E_{out}** del modelo lo más ajustada posible.
10. **Discutir y justificar** la calidad del modelo encontrado y las razones por las que considera que dicho modelo es un buen ajuste que representa adecuadamente los datos muestrales.

Ajuste de Modelos Lineales

- En la práctica:
 - Solamente se pide emplear **modelos lineales** (regresión lineal, regresión logística y perceptrón+pocket), junto con las **transformaciones en los datos**, técnicas de **regularización** y **preprocesado** que consideréis más conveniente
 - Si alguien quiere probar a mayores SVM, MLP, RF. ¡Perfecto! Que compare con los modelos lineales y justifique su uso. ¡Pero hay que usar modelos lineales!

Ajuste de Modelos Lineales

.zip = Código (.py) + Informe (.pdf)

Fecha de entrega: 20 de Mayo

SE VALORARÁ ENORMEMENTE LA **JUSTIFICACIÓN** DE LAS DECISIONES TOMADAS Y LA **DISCUSIÓN** DE LOS RESULTADOS OBTENIDOS.