

Prácticas de Aprendizaje Automático Grupo 1

Trabajo 3: Ajuste de Modelos Lineales

Francisco Javier Baldán Lozano
Universidad de Granada

Departamento de Ciencias de la Computación e Inteligencia Artificial



UNIVERSIDAD
DE GRANADA



Recordatorio normas (1). Informe.

.zip = Código (.py) + Informe (.pdf)

- Presentar un **informe escrito** con las valoraciones y **decisiones** adoptadas en cada apartado.
 - No es solo hacer algo → hay que argumentar el por qué
- Incluir en el informe los **gráficos** generados.
- Incluir una **valoración/discusión de los resultados obtenidos.**
- **El informe debe presentarse en PDF**
- Si no hay informe → se considera que el trabajo no se ha presentado.

Recordatorio normas (2). Código.

- **Un script de Python por ejercicio.**
 - Los distintos ejercicios van en diferentes ficheros .py.
- **Todos los resultados numéricos o gráficas serán mostrados por pantalla**, parando la ejecución después de cada apartado.
 - No escribir nada en el disco.
- El path que se use en la lectura de cualquier fichero auxiliar de datos debe ser siempre **"datos/nombre_fichero"**.
 - Crear directorio llamado "datos" dentro del directorio donde se desarrolla y se ejecuta la práctica.

Recordatorio normas (3). Código.

- El código **debe ejecutarse de principio a fin sin errores.**
- No es válido usar opciones en las entradas.
 - **Fijar al comienzo los parámetros por defecto** que considere óptimos.
- **El código debe estar obligatoriamente comentado** explicando lo que realizan los distintos apartados.
 - **Id comentando el código** que hagáis: sirve para que entendáis mejor lo que habéis hecho, y facilita mi trabajo a la hora de corregir los ejercicios.
- **Entregar solo el código fuente, nunca los datos.**

Recordatorio normas (y 4)

.zip = Código (.py) + Informe (.pdf)

Subir el zip en la correspondiente entrega en PRADO.

Fecha de entrega: 30 de Mayo

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

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
 - <http://archive.ics.uci.edu/ml/datasets/Communities+and+Crime>



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:	268110



Communities and Crime Data Set

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

Abstract: Communities within the United States. The data combines socio-economic data from the 1990 US Census, law

Data Set Characteristics:	Multivariate	Number of Instances:	1994	Area:	Social
Attribute Characteristics:	Real	Number of Attributes:	128	Date Donated	2009-07-13
Associated Tasks:	Regression	Missing Values?	Yes	Number of Web Hits:	278205

Ajuste de Modelos Lineales

- 1) (1 Punto) Comprender el problema a resolver → ¡Visualizar datos!
- 2) (1.5 Puntos) Selección de clases de **funciones a usar** → combinaciones lineales/cuadráticas/etc. de los valores observados
- 3) (1.5 Puntos) Definición de los conjuntos de **training, validación y test** usados en su caso → ¿uso de cross-validation?
- 4) (2.5 Puntos) Preprocesado los datos: codificación, normalización, proyección, etc. Es decir, todas las manipulaciones sobre los datos iniciales hasta fijar el conjunto de vectores de características que se usarán en el entrenamiento.
- 5) (1.5 Puntos) Fijar y discutir la idoneidad de la **métrica** usada en el ajuste → MSE, MAE, Cross-Entropy,...
- 6) (1.5 Puntos) Discutir la técnica de ajuste elegida.
- 7) (2.5 Puntos) Discutir la necesidad de regularización y, en su caso, la función usada para ello.
- 8) (2.5 Puntos) Definir los modelos **LINEALES** a usar.
- 9) (1 Punto) Estimación de hiperparámetros y selección del mejor modelo
- 10) (1.5 Puntos) Estimación por validación cruzada del error E_{out} del modelo. Compárela con E_{test} , ¿qué conclusiones obtiene?
- 11) (3 Puntos) Suponga que Ud ha sido encargado de realizar este ajuste para una empresa. ¿Qué modelo les propondría y que error E_{out} les diría que tiene?. Justifique las decisiones

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!

Bibliografía

Transparencias realizadas y actualizadas a partir del trabajo de Pablo Mesejo.