

Práctica 4. Regresión Logística Multi-Clase y R.N.

Objetivo

El objetivo es resolver mediante regresión logística un problema real de clasificación multi-clase: el reconocimiento de dígitos manuscritos. Utilizaremos el conjunto de datos MNIST. Cada muestra es una imagen de 28x28 píxeles.

Estudio previo

- A. Repasa las transparencias de teoría y estudia a fondo las funciones de Scikit-Learn que se indican en la última transparencia del tema.

Desarrollo de la práctica

El programa `P4demo.py` carga los datos de MNIST. Como verás, los datos ya están normalizados, y permutados aleatoriamente. A continuación, escribe los programas necesarios para **entrenar, evaluar y comparar entre sí distintos modelos** de clasificación multiclase por regresión logística y redes neuronales, siguiendo los siguientes pasos:

1. **Curvas de aprendizaje.** Dibuja las curvas de aprendizaje de un clasificador por regresión logística sin regularizar, variando en número de muestras de entrenamiento entre el 10% y el 100% de las disponibles. Analiza los resultados y extrae conclusiones.
2. **Regresión logística.** Entrena un modelo utilizando regresión logística, que clasifique los caracteres lo mejor posible. Evalúa con los datos de test el mejor modelo que hayas obtenido. Analiza qué dígitos son más problemáticos, y qué confusiones son más habituales.
3. **Perceptrón sin capa oculta.** Entrena una red sin capa oculta, para la clasificación multiclase. Evalúa con los datos de test el mejor modelo obtenido, analiza y discute los resultados.
4. **Perceptrón multicapa.** Busca un modelo que utilice un perceptrón multicapa, que clasifique los caracteres lo mejor posible. Evalúa con los datos de test el mejor modelo que hayas obtenido. Analiza qué dígitos son más problemáticos, y qué confusiones son más habituales.
5. Escribe en la memoria un **resumen de las conclusiones** principales obtenidas. Almacena en un dataframe de pandas las métricas de los mejores modelos encontrados en los apartados anteriores, y sálvalo en un fichero para poder compararlos con los resultados de la siguiente práctica.

A entregar en Moodle

- Un notebook `P4.ipynb` con los programas de cada apartado, los resultados, su interpretación y las conclusiones que hayas obtenido.
- O, si lo prefieres, un fichero `P4.py` con el código y otro `P4.pdf` con los resultados de todos los apartados, su interpretación y las conclusiones que hayas obtenido.

Recuerda:

- Trae la práctica preparada para aprovechar la sesión de prácticas al máximo.
- Si te atascas, pregunta en la sesión o en tutorías.
- Debes citar correctamente todas las fuentes utilizadas.
- Tienes 5 días desde tu sesión para depositar la práctica en Moodle.
- Deberás defenderla en tu próxima sesión de prácticas.