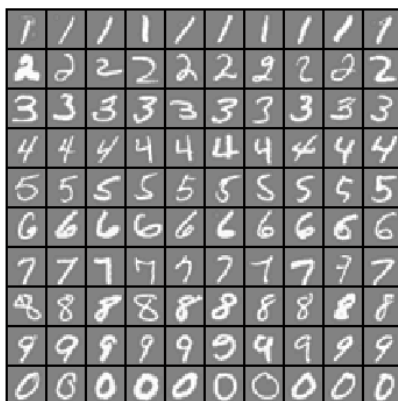


Práctica 4. Regresión Logística Multi-Clase

Objetivo

El objetivo es resolver mediante regresión logística un problema real de clasificación multi-clase: el reconocimiento de dígitos manuscritos. Utilizaremos una versión reducida del conjunto de datos MNIST <http://yann.lecun.com/exdb/mnist/>. La figura muestra un ejemplo de los mismos. Cada muestra es una imagen de 28x28 píxeles. Como atributos para la clasificación utilizaremos directamente los niveles de intensidad de los 784 píxeles.



Estudio previo (es necesario prepararlo **por escrito, antes** de acudir a la práctica)

1. Repasa las transparencias de clase y estudia las funciones auxiliares proporcionadas para esta práctica. Escribe el algoritmo de entrenamiento y clasificación multi-clase utilizando regresión logística regularizada.

Desarrollo de la práctica

Copia a tu directorio de trabajo los ficheros proporcionados, y comprueba que funcionan correctamente en Matlab. Comprueba también la función `minFunc`. A continuación escribe los programas necesarios para resolver la regresión logística, siguiendo los siguientes pasos:

2. **Regresión logística regularizada.** Basándote en el código de la práctica anterior, programa el entrenamiento y clasificación multi-clase. Separa un 20% de los datos para validación, y resuelve la regresión logística regularizada, dibujando la gráfica de tasa de errores en función del parámetro λ .
3. **Matriz de confusión y Precisión/Recall.** Re-entrena con todos los datos para el mejor valor de λ , y utiliza los datos de test para calcular la matriz de confusión y los valores de precisión y recall para cada dígito, y visualiza las confusiones con la función `verConfusiones`. ¿Qué dígitos son los más problemáticos?

A entregar (en Moodle, dentro de un fichero .zip)

- Programa `P4.m`, junto con las funciones auxiliares que hayas programado, que vaya mostrando por pantalla los resultados de todos los apartados.
- Si no presentas la práctica durante la sesión, además deberás entregar la memoria de la práctica en un fichero `P4.pdf` ó `P4.doc` con los resultados de todos los apartados, su interpretación y las conclusiones que hayas obtenido.