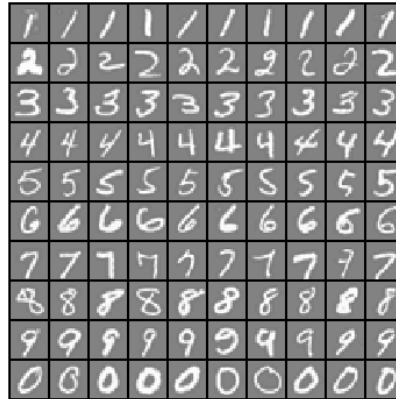


## Práctica 4. Regresión Logística Multi-Clase

### 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)

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

### Desarrollo de la práctica

Copia a tu directorio de trabajo los ficheros proporcionados, y comprueba que funcionan correctamente en Matlab. Atención, la instrucción `load('MNISTdata2.mat')` carga tanto las variables `X` e `y` que se usan para el entrenamiento, como `Xtest` e `ytest` que sólo se usarán para la evaluación final del modelo. A continuación, escribe los programas necesarios para resolver la regresión logística, siguiendo los siguientes pasos:

1. **Regresión logística regularizada.** Basándote en el código de la práctica anterior, programa el entrenamiento y clasificación multi-clase one-vs-all. Entrena la regresión logística regularizada, buscando el mejor valor de regularización  $\lambda$ , utilizando para ello la métrica que te parezca más adecuada para este problema. Dibuja la evolución de la métrica elegida en función de  $\lambda$ , y analiza el comportamiento del modelo.
2. **Evaluación del modelo final.**
  - a. Re-entrena el mejor modelo obtenido con todos los datos de entrenamiento, y evalúalo.
  - b. Analiza qué dígitos son más problemáticos, y qué confusiones son más habituales.

**A entregar en Moodle**

- Memoria de la práctica en un fichero `P4.pdf` con el escaneado del estudio previo, los resultados de todos los apartados, su interpretación y las conclusiones que hayas obtenido.
- Programa `P4.m` que vaya ejecutando cada apartado y mostrando por pantalla los resultados obtenidos, junto con las funciones auxiliares que hayas programado, en un fichero `.zip`

**Recuerda que:**

- Tienes 5 días desde tu sesión de prácticas para depositarla en Moodle.
- Deberás defender la práctica en tu próxima sesión de prácticas.