Práctica 3. Regresión Logística

Objetivo
El objetivo es aplicar regresión logística en casos sencillos de clasificación binaria, utilizando técnicas de regularización y validación cruzada.

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, incluida la funcion de optimización minFunc con sus diferentes opciones. Escribe el algoritmo de k-fold cross-validation para elegir el valor del parámetro de regularización.

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 básica**. Queremos predecir qué alumnos serán admitidos a una universidad, en función de la calificación obtenida en dos exámenes, aprendiendo a partir de los datos del fichero exam_data.m. Separa un 20% de los datos para test. Programa la función de coste y resuelve la regresión logística. Calcula la tasa de errores con los datos de entrenamiento y con los datos de test. Para un alumno que ha sacado 45 puntos en el primer examen, dibuja una gráfica con la probabilidad de ser admitido en función de la calificación del segundo examen.

3. **Regularización.** En el control de calidad de una planta de fabricación cada microchip pasa una serie de test de funcionamiento, y queremos decidir si aceptarlos o rechazarlos utilizando únicamente los resultados de dos tests. Para ello contamos con un conjunto de datos mchip_data.txt con los resultados obtenidos por un conjunto de microchips y si fueron finalmente aceptados o rechazados. Separa un 20% de los datos para test. Utilizaremos regresión logística regularizada con expansión de funciones base mediante la función mapFeature.m proporcionada. Elige el parámetro de regularización landa mediante k-fold cross-validation. Dibuja las curvas de evolución de las tasas de errores con los datos de entrenamiento y de validación. Finalmente, entrena con todos los datos (excepto los de test) el mejor modelo encontrado y el modelo con landa=0, y dibuja las correspondientes superficies de separación. ¿Cuál de los dos modelos es mejor?

4. **Precisión/Recall.** Con el mejor modelo obtenido, utiliza los datos de test para calcular la matriz de confusión y los valores de precisión y recall. Si queremos que el 95% de los chips aceptados sean buenos, ¿Qué habría que hacer?

A entregar (en Moodle, dentro de un fichero .zip)
- Programa `P3.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 `P3.pdf` ó `P3.doc` con los resultados de todos los apartados, su interpretación y las conclusiones que hayas obtenido.