

Tarea programada 4

Curso: Inteligencia Artificial

30 de marzo de 2017

Instituto Tecnológico de Costa Rica
Sede Interuniversitaria de Alajuela
Prof. Fabián Fallas Moya

Introducción

En esta tarea usted va a implementar regresión logística y ver cómo funciona sobre datos. Suponga que usted es el administrador del departamento de admisiones de una universidad y quiere saber las posibilidades de que los estudiantes sean admitidos, basados en el resultado de dos exámenes. Usted cuenta con registros históricos de estudiantes anteriores, por cada estudiante usted posee el resultado de las pruebas. Su trabajo es crear un modelo de clasificación basado en el resultado de las dos pruebas e indicar las posibilidades de que un estudiante sea admitido o no. El archivo `hw4.m` lo guiará para llevar a cabo esta tarea.

Archivos para la tarea (templates)

A través de esta tarea usted va a utilizar los scripts `hw4.m` y `hw4_reg.m`. Estos archivos van a inicializar y cargar los datasets y hacer las llamadas a las funciones necesarias para ejecutar todos los ejercicios. No es necesario modificar estos archivos, lo que sí es necesario es modificar las funciones que se encuentran en otros archivos, y seguir las instrucciones que se especifican en este documento.

1.1 hw4 y visualización

Lo primero que deberá completar es el archivo para graficar adecuadamente estos datos (archivo `plotData.m`). En el gráfico podremos observar todos los estudiantes anteriores y si fueron admitidos o no.

1.2 Función sigmoide

En esta parte del ejercicio usted deberá completar el código para ejecutar la función sigmoide (tome en cuenta que z puede ser vector, matriz o un escalar).

1.3 Función de costo y gradiente

Ahora debe completar el archivo `funcionCosto.m`. Al final, el costo inicial (significa que con los parámetros θ en cero) tendrá un valor de 0.693 aprox.

1.4 fminunc

En esta sección usted va a llamar a `fminunc` (ya está el código), si todo está bien el costo debe ser aprox. de 0.203

1.5 Evaluando

Con el estudiante cuyas notas son 45 y 85, se debería de apreciar una probabilidad de admisión de 0.776. Al final del ejercicio usted podrá apreciar la exactitud del modelo con respecto al set de entrenamiento.

2.1 hw4_reg

En esta parte del ejercicio usted implementará regresión logística regularizada para predecir si los microchips de una planta de fabricación pasan las pruebas de calidad - Quality Assurance (QA). Durante QA, cada microchip pasa por varias pruebas para asegurarnos si funciona bien.

Suponga que usted es el administrador de una fábrica y desea ver con los resultados de dos diferentes pruebas, si los chips deben ser aceptados o rechazados.

2.2 Visualizando

Si usted logra visualizar los datos, podrá darse cuenta de que no son linealmente separables (significa que con una línea no podremos separar o clasificar).

2.3 Mapeo de Características

En el archivo de mapeo de características no debe modificar nada, simplemente entender qué es lo que está sucediendo (se están creando muchas características a partir de las únicas dos características con las que contamos). Pasamos de 2 a un vector de dimensión 28.

2.4 Función de costo y gradiente

Ahora lo que debe hacer es completar el código de `funcionCostoReg.m` para retornar el costo y el gradiente. Haga uso de las fórmulas vistas en clase para completar este apartado (el costo inicial debe ser de 0.693).

Al finalizar el ejercicio debe de poder visualizar un decision boundary adecuado para los datos.