

Inteligencia Artificial - Laboratorio 4 -

Instrucciones:

- Esta actividad debe realizarse individualmente.
- No se permitirá ni se aceptará cualquier indicio de copia. De presentarse, se procederá según el reglamento correspondiente.
- Tendrán hasta la fecha indicada en Canvas para entregar este laboratorio.
 - No se confíen, aprovechen el tiempo en clase para entender todos los ejercicios y avanzar lo más posible.

Task 1 - Regresión logística polinómica

Instrucciones: Usted usará Python a través de un Jupyter Notebook para llevar a cabo este ejercicio. Recuerde utilizar comentarios para describir lo que está haciendo en cada parte del proceso. Estará usando el juego de datos proporcionado dentro del portal. Al finalizar recuerde subir al portal un link a su repositorio en el que se pueda correr su notebook, usando <https://mybinder.org/>.

El juego de datos proporcionado es parte de la plataforma Kaggle, dentro del cual se muestran condiciones físicas y contextuales para más de 4000 pacientes de enfermedades cardíacas. El dataset relaciona a cada paciente con una etiqueta (1 = tuvo un paro cardíaco, 0 = no tuvo paro cardíaco).

Abajo una breve descripción de las variables incluidas:

- **Demographic:**
 - Sex: male or female(Nominal)
 - Age: Age of the patient,(Continuous - Although the recorded ages have been truncated to whole numbers, the concept of age is continuous)
- **Behavioral:**
 - Current Smoker: whether or not the patient is a current smoker (Nominal)
 - Cigs Per Day: the number of cigarettes that the person smoked on average in one day.(can be considered continuous as one can have any number of cigarettes, even half a cigarette.)
 - Medical(history)
 - BP Meds: whether or not the patient was on blood pressure medication (Nominal)
 - Prevalent Stroke: whether or not the patient had previously had a stroke (Nominal)
 - Prevalent Hyp: whether or not the patient was hypertensive (Nominal)
 - Diabetes: whether or not the patient had diabetes (Nominal)
- **Medical(current):**
 - Tot Chol: total cholesterol level (Continuous)
 - Sys BP: systolic blood pressure (Continuous)
 - Dia BP: diastolic blood pressure (Continuous)
 - BMI: Body Mass Index (Continuous)
 - Heart Rate: heart rate (Continuous - In medical research, variables such as heart rate though in fact discrete, yet are considered continuous because of large number of possible values.)
 - Glucose: glucose level (Continuous)

Para este ejercicio se le pide que proporcione un modelo de regresión logística polinomial que prediga fielmente si un paciente sufrirá de un paro cardíaco.

Task 1.1

Leer el archivo CSV proporcionado y almacenarlo en un np.array para ser trabajado en el notebook.

Inteligencia Artificial - Laboratorio 4 -

Task 1.2

Ajustar un modelo logístico polinomial en base al juego de datos cargado de forma matricial que relaciona las variables independientes que usted considere apropiadas (puede no utilizar todas las componentes de X), con la variable dependiente de salida (sufrir o no sufrir un paro cardíaco)..

Task 1.3

Utilice la implementación vectorial del algoritmo de regresión logística (descenso del gradiente visto en clase).

Task 1.4

Usando cross-validation determine el grado del polinomio que mejor describe la nube de puntos (encuentre el mejor balance entre apego a los datos de entrenamiento y generalización para datos previamente no observados).

Task 1.5

Haga un análisis sobre sus hallazgos.

Entregas en Canvas

1. Link al repositorio.
 - a. Deberán subir el código también a Canvas por temas de Acreditación

Evaluación

1. [0.5 pts.] Task 1.1
2. [0.5 pts.] Task 1.2
3. [1 pts.] Task 1.3
4. [1.5 pts.] Task 1.4
5. [1.5 pts.] Task 1.5

Total 5 pts.