

Universidad Autónoma de Baja California

Facultad de Ciencias Químicas e Ingeniería

Ingeniería en Computación

Evaluación Meta 5.5, Aplicaciones: Inteligencia Artificial

Fecha: 27/Mayo/2022

Nombre Integrantes del Equipo: _____ Cal: _____

1. Diseñar modelos de regresión a partir de los datos (bodyfat_dataset.dat) que establecen la relación entre la masa corporal (targets, última columna de la tabla) de 252 pacientes a partir de 13 medidas predictoras (inputs, primeras 13 columnas de la tabla) : Edad (años), Peso (libras), Altura (pulgadas), Circunferencia del cuello (cm), Circunferencia del pecho (cm), Circunferencia del abdomen 2 (cm), Circunferencia de la cadera (cm), Circunferencia del muslo (cm), Circunferencia de la rodilla (cm), Circunferencia del tobillo (cm), Circunferencia de bíceps (extendida) (cm), Circunferencia del antebrazo (cm) y Circunferencia de la muñeca (cm).

El objetivo del análisis de regresión es una forma de técnica de modelado predictivo que investiga la relación entre una variable dependiente (objetivo) y una variable independiente (predictor). Esta técnica se utiliza para pronosticar y encontrar la relación del efecto causal entre las variables.

2. Diseñar modelos de regresión a partir de los datos (reaction_dataset.dat) que establecen la relación entre la velocidad de reacción (targets, última columna de la tabla) y concentraciones de tres reactivos como medidas predictoras (inputs, primeras tres columnas de la tabla): x_1 (hidrógeno), x_2 (n-pentano), x_3 (isopentano). La velocidad de reacción (\hat{y}) para la cinética de reacción es el modelo de Hougen-Watson:

$$\hat{y} = \frac{\theta_1 x_2 - \frac{x_3}{\theta_5}}{1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3}$$

donde \hat{y} es la velocidad de reacción estimada, x_1 , x_2 , x_3 son las concentraciones de hidrógeno, n-pentano e isopentano, respectivamente, y θ_1 , θ_2 , ..., θ_5 son parámetros del modelo.

3. Diseñar modelos de clasificación logística para detectar cáncer a partir de datos de espectrometría de masas en perfiles de proteínas (ovarian_dataset.dat) que establecen la relación para distinguir entre pacientes con cáncer (targets, últimas dos columnas de la tabla) y medidas de intensidad de iones (inputs, primeras 100 columnas de la tabla).

La metodología es seleccionar un conjunto reducido de medidas o "características" que pueden usarse para distinguir entre pacientes con cáncer y de control mediante un clasificador. Estas características serán niveles de intensidad de iones a valores específicos de masa / carga. El objetivo

del modelo de clasificación es distinguir entre el cáncer y el control de los pacientes a partir de los datos de espectrometría de masas.

4. Diseñar modelos de clasificación logística para detectar enfermedades de la piel (dermatology_dataset.dat) de 366 pacientes, que establecen la relación para clasificar las enfermedades de psoriasis, dermatitis seborreica, liquen plano, pitiriasis rosada, dermatitis crónica y la pitiriasis rubra pilaris (targets, última columna de la tabla) y medidas con 12 características clínicas mas 22 características histopatológicas (inputs, primeras 34 columnas de la tabla).

Fecha de entrega: 2 días antes del examen Ordinario.