

Inteligencia Artificial
Proyecto #1: Clasificación
Prof. Cristian López Del Alamo

Este conjunto de datos contiene imágenes de diez categorías (especies) de mariposas. El conjunto de datos de imágenes consta de 832 imágenes en total, con una distribución que varía de 55 a 100 imágenes por categoría.

El nombre de los archivos indican la especie y la imagen correlativa de esa especie. Por ejemplo, el archivo 0070038, indica que es la imagen número 38 de la especie 7. Tenga esto en cuenta para poder separar las especies y las imágenes.

Actividades

1. Utilizar el *dataset* **Mariposas**.
2. Para cada imagen genere vectores característicos, estos pueden ser generados por medio de **hard wavelet** o buscar algunos otros métodos para la extracción de características que su grupo pueda implementar.
3. Implemente los algoritmos de clasificación Regresión Logística, SVM, KNN y Árboles de decisión. (Si utiliza librerías la nota máxima será sobre 15).
4. Tanto Regresión logística como SVM son modelos para clasificación binaria, pero pueden ser extendidos para funcionar para n clases. Su equipo deberá ver la forma de aplicarlo a varias clases.
5. Separe el dataset en 3 grupos. (70 % para entrenamiento, 15 % para validación y 15 % para testing).
6. Entrene los modelos usando *K-fold cross validation* o *Bootstrap*.
7. Genere las gráficas de pérdida vs épocas para analizar el comportamiento del entrenamiento y validación.
8. En una tabla, coloque los valores de Precisión, *Recall*, *F1-Score* hiper-parámetros en cada modelo.
9. Finalmente, concluya que modelos le entregan los mejores resultados.

NOTA: Cada equipo deberá subir un solo documento con la siguiente estructura:

1. Introducción.
2. Explicación de los modelos así como las consideraciones tomadas en cuenta para generar los vectores característicos.
3. Experimentos. Es la parte más importante del proyecto y debe ser realizado de manera exhaustiva. Utilice gráficos y tablas mostrar sus hallazgos.
4. No olvide hacer las pruebas variando los hiper-parámetros del modelo.
5. Conclusiones: Redacte las conclusiones de acuerdo a los resultados.

Sobre la Implementación:

Cada equipo es responsable del funcionamiento de su código. Para ser evaluado correctamente su código debe ser funcional; es decir, debe permitir ejecutar todas las pruebas llamando a una única función que genere todas las salidas, tal cual, se encuentren consignadas en su documento.

Finalmente, anexar el link del github o el colab para verificar el código fuente. Así mismo, definir una **semilla** para replicar los resultados.

Sobre copias:

Toda copia parcial o total invalida el trabajo y generará activar el protocolo de copia ante las autoridades de UTEC.