



Práctico Obligatorio Machine Learning

1 Objetivo

El objetivo de la presente práctica es adquirir conocimientos básicos sobre la utilización del aprendizaje automático implementando uno de los métodos de aprendizaje y comparando los resultados con el paquete WEKA.

2 Herramientas

2.1 WEKA



WEKA (Waikato Environment for Knowledge Analysis - Entorno para Análisis del Conocimiento de la Universidad de Waikato) es un conocido software para aprendizaje automático y minería de datos escrito en Java y desarrollado en la Universidad de Waikato. WEKA es un software libre distribuido bajo licencia GNU-GPL.

URL oficial de Weka: https://waikato.github.io/weka-wiki/downloading_weka/

3 Asignación de grupos

Asignación de algoritmos:

Grupo 1 Lara Acuña, Giovanni Sicolo, y Diego Baltar.

Tema: Árbol de decisión (elegir: ID3, C4.5 u otro).

Grupo 2 Antonella Torres y Adrian Lusso.

Tema: Red neuronal.

4 Data-sets

Los data-sets a utilizar se pueden descargar de las siguientes páginas Webs:

Wine: <https://archive.ics.uci.edu/dataset/109/wine>

Zoo: <https://archive.ics.uci.edu/dataset/111/zoo>

Algerian Forest Fires: <https://archive.ics.uci.edu/dataset/547/algerian+forest+fires+dataset>

5 Ejercicios

Ejercicio 1: Investigar el método de aprendizaje automático asignado. Su funcionamiento, características y su implementación.

Ejercicio 2: Seleccionar un data-set de los indicados previamente. Clasificar los datos para entrenamiento y para testing del método de aprendizaje.

El set de entrenamiento debe ser el 80% del set de datos. El set de test deben ser instancias aleatorias pero bien distribuidas con respecto al atributo a clasificar.

Ejercicio 3: Implementación

- Implementar en el lenguaje de preferencia la construcción del algoritmo asignado.
- Construir un modelo (árbol, red neuronal, hiperplano) por set de entrenamiento a partir del algoritmo implementado.
- Evaluar el modelo construido (árbol, red, hiperplano) con el set de test armado en el ejercicio previo. Obtener la matriz de confusión.

No se permite el uso de librerías de terceros que implementen parte o la totalidad del algoritmo indicado en el enunciado, o un derivado del mismo. La implementación del algoritmo debe ser enteramente propia de los integrantes del grupo. Las librerías ya existentes y proveídas por otras personas no serán consideradas como parte de la resolución del enunciado.

Ejercicio 4: WEKA

Comparar los resultados encontrados en el punto anterior con el mismo set de datos sobre WEKA.

1. Crear el modelo del método elegido (árbol o red neuronal) utilizando el mismo set de entrenamiento.
2. Utilizando el mismo set de test del ejercicio anterior, evaluar el modelo obteniendo la matriz de confusión.

Ejercicio 5: Análisis y evaluación de los modelos.

Comparar y analizar todos los resultados obtenidos de su implementación con respecto a los resultados dados por Weka.

Ejercicio 6: Escribir un informe como se indica en las siguientes secciones.

Ejercicio 7: Realizar una presentación de su tema al resto de la clase explicando cómo funciona el algoritmo, el modelo realiado y el análisis y comparación de los resultados obtenidos. El tiempo de la presentación será indicado por la cátedra oportunamente.

6 Consideraciones

El trabajo se puede realizar en grupos de no más de 3 estudiantes.

El informe será entregado en forma digital en formato PDF en un formulario de Google de la URL especificada más abajo. Se debe poder reproducir la experimentación realizada en todos los ejercicios.

Comprimir los archivos en un zip/tar con los apellidos de los integrantes del grupo.

7 Informe - Presentación

El informe para este trabajo práctico es obligatorio y debe ser acompañado de una exposición ante el resto de la clase. El informe deberá ser de más de 5 hasta 15 páginas. Se sugiere la siguiente estructura:

1. Carátula:

Identifique claramente los integrantes del grupo, legajo y dirección de email. Se recomienda indicar las carreras en la que se inscribieron a esta materia.

2. Introducción (motivación).

3. Dominio y explicación del data-set asignado.

4. Marco teórico: Análisis y explicación del algoritmo asignado.

5. Características de la implementación:

Implementación realizada y cómo se ejecuta.

6. Resultados de la ejecución

- Análisis de las corridas.
- Análisis de las corridas con WEKA.

7. Comparación, análisis y discusión.

8. Conclusiones y posibles mejoras.

9. Bibliografía y/o referencias



Formulario de Google de entrega:

<https://forms.gle/6zY6LfdVTwguPuW69>



Fecha de primer entrega borrador: Miércoles 12 de junio a las 23:59.

Esta entrega no es definitiva. Es para resolver consultas y observar el avance del informe.



Fecha límite de recepción del informe: 19 de junio a las 23:59.



Fecha de la presentación: Tentantiva - Lunes 24 de junio a las 15:00.