

Curso de Postgrado:
“Aprendizaje Automático: Fundamentos y Aplicaciones en Meteorología del Espacio”

DCAO-UBA - Agosto 2022

Dra. María Graciela Molina (FACET-UNT / CONICET)

Lic. Jorge H. Namour (FACET-UNT /)

Github: <http://shorturl.at/jWX46>

Trabajo práctico N° 1: Clasificación

Problema N° 1:

Existen diferentes métodos de monitoreo de la ionosfera, uno de los más usados es con sistemas de radar que aprovechan las características reflectivas de la ionósfera y las propiedades de la propagación de ondas electromagnéticas. Un problema que se puede aplicar a diferentes sistemas de radar tiene que ver con la capacidad de clasificar los ecos recepcionados en la antena del radar. En este ejercicio se propone analizar un set de datos correspondientes a ecos recibidos. Cuando los mismos son el resultado de una reflexión en la ionósfera, se debe clasificar como un eco válido. Se dispone del archivo “ionosphere.data” con ecos registrados que ya se encuentran clasificados (retorno positivo o no) manualmente por expertos.

Para este problema:

1. Se desea armar un clasificador que permita determinar, a partir de los ecos (features) disponibles del radar si se trata de un eco positivo o no (función target).
2. Durante la etapa de ingeniería de datos:
 - a. Determinar si los datos están balanceados.
 - b. Proponer alguna técnica para su balanceo en caso de que exista un desbalanceo significativo.
3. Analizar los resultados a partir de varios modelos generados.
4. Mostrar gráfica de loss function y accuracy de cada uno de los modelos.

Problema N° 2:

Las eyecciones de masa coronal o *Coronal Mass Ejections* (CMEs) son uno de los eventos más importantes que ocurren en el Sol. En muchos casos el origen de las CMEs pueden ser las denominadas fulguraciones solares o “flares”, aunque existen otros eventos que dan origen a las CMEs (e.g. filamentos). El archivo “CMEs.csv” contiene un listado de CMEs catalogadas según su origen (flare/no-flare), y además cuenta con dos features correspondientes a la actividad solar en el momento de ocurrencia de la CME (baja/alta) y a la velocidad de propagación. Las columnas que posee el archivo son: “SunActivity”, “vprop” y “Origin”.

- “SunActivity”: variable categórica que indica actividad solar en el año que se tomó el registro (low, high).
- “Vprop”: variable numérica que indica la velocidad del viento solar durante el registro.
- “Origin”: variable binaria que indica si hubo o no hubo un flare (flare, no flare) antes de una CME.

Curso de Postgrado:

“Aprendizaje Automático: Fundamentos y Aplicaciones en Meteorología del Espacio”

DCAO-UBA - Agosto 2022

Dra. María Graciela Molina (FACET-UNT / CONICET)

Lic. Jorge H. Namour (FACET-UNT /)

Github: <http://shorturl.at/jWX46>

Para este problema:

1. Se desea armar un clasificador sobre la variable “Origin”.
2. Determinar si los datos están balanceados.
3. ¿Qué tipo de categorización de variables es necesaria?
4. Respecto a los modelos generados:
 - a. ¿Qué problemas se pueden identificar en los datos que perjudican al entrenamiento?
 - b. ¿Poseen una precisión aceptable según su punto de vista?
 - c. ¿Qué alternativas propondría para mejorar los modelos?
5. Analizar los resultados a partir de varios modelos generados.
6. Mostrar gráfica de loss function y accuracy de cada uno de los modelos.