

Fecha de liberación oficial: martes 26 de marzo 23:01

## **Tarea para el Hogar DOS**

Todo lo que estamos haciendo en esta asignatura es por dinero; queremos lograr un modelo predictivo con mayor ganancia posible en pesos argentinos, realizamos experimentos para entender que algoritmos y sus parámetros que debemos utilizar, en que datos debemos entrenar, que tipo de Feature Engineering debemos generar. A veces, las mejoras provienen de utilizar un método de procesamiento más veloz, para disponer de más tiempo para nuevos experimentos.

Aclaración importante : algunos experimentos que realizaremos nos darán menos ganancia es parte de la exploración.

Todo lo que usted correrá en esta Tarea para el Hogar será en Google Cloud, la mayoría de las veces en la máquina virtual [desktop](#) que reside en Sao Paulo

## **Sección Pasado**

### **1. CheckPoint**

A este momento usted ya debe haber completado:

1. Arranque en Frío, el capítulo 3 de *El Libro de la Asignatura*
2. Ha realizado varios submits a Kaggle
3. Instalación de Google Cloud, disponible en la *Tarea para el Hogar 1UNO*
4. Tiene disponible en Google Cloud la máquina virtual llamada [desktop](#) que reside en Sao Paulo y ha ingresado a ella siguiendo la demo realizada durante la segunda clase.

## Sección Deseable

### 2. Lecturas Faltantes

de *El Libro de la Asignatura* lea

- capítulo 4 Herramientas, Conceptos, Operación y Buenas Prácticas
- capítulo 5 Plataforma Kaggle

(tiempo estimado 20 minutos, , dificultad media)

### 3. Zero2Hero Primera Parte

Se han lanzado los primeros fascículos coleccionables llamados "from Zero to Hero" que muy detalladamente, paso a paso enseñan todo lo necesario de R para entender los scripts oficiales de la asignatura.

Están en formato Jupyter Notebook, para ejecutarlos paso a paso utilice Jupyter Lab o directamente VSCode desde la máquina virtual [desktop](#) que reside en Sao Paulo

Podrá encontrar los Jupyter Notebooks zero2hero en el repositorio GitHub de la asignatura carpeta [./src/zero2hero](#) vea los fascículos del zero2hero\_0101 al zero2hero\_0112 inclusive

Si no sabe programar, esta es su oportunidad de empezar a aprender :  
ejecute los notebooks línea a línea

[pregunte a ChatGPT que hace cada línea que no entienda](#)

experimente cambiar instrucciones y ver que pasa, recuerde siempre hacer una copia del script quitándole la "z" inicial.

Utilice Zulip para preguntar/comentar en el stream [# zero2hero](#)

(tiempo estimado 60 minutos, , dificultad media)

### 4. Optimización Hiperparámetros

Si aún no lo ha hecho ver el video correspondiente a la Clase2 que están en en campus

- [Optimización de Hiperparámetros en un Arbol de Decisión](#)  
( tiempo estimado 9 minutos a 1.5x, dificultad media)

## 5. Corrida Grid Search

Modifique el script [src/rpart/z141\\_gridsearch\\_esqueleto.r](#) agregando loops para que recorra TODOS los hiperparámetros de rpart < cp, maxdepth, minsplit, minbucket >, y luego póngalo a correr.

Tenga muy presente la granularidad que eligirá para cada hiperparámetro.

En la línea 14, reemplace por sus propias semillas

```
ksemillas <- c(102191, 200177, 410551, 552581, 892237) #reemplazar por las propias semillas
```

dificultad media-alta

tiempo promedio

- trabajo humano : 30 minutos
- procesamiento : 4 horas

## 6. Análisis de resultados de Grid Search

- Levante el archivo de salida de la corrida anterior en una planilla tipo Excel
- Ordene por `ganancia_promedio` descendente
- El de mayor `ganancia_promedio` decimos que es el primero del ranking
- Cargue los resultados de los rankings 1, 2, 5, 10, 50 y 100 en la hoja **GridSearch** de la Planilla Colaborativa de la Asignatura ( agregue su nombre a un formulario vacío )  
[https://docs.google.com/spreadsheets/d/1SXAr21laWIV3Agk6QiWuaS8B9JsNc\\_7DZe56XLNuY94/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1SXAr21laWIV3Agk6QiWuaS8B9JsNc_7DZe56XLNuY94/edit?usp=sharing)
- Para cada una de esas seis combinaciones de hiperparámetros, utilizando el script `src/rpart/z101_PrimerModelo.R` genere un submit, súbalo a Kaggle, y copie la ganancia del Public Leaderboard en la correspondiente celda de la planilla
- ¿Se cumple que la mejor de Estimación Montecarlo es la mejor ganancia del Public Leaderboard?
- Si lo anterior no sucede : [https://www.youtube.com/watch?v=Bti9\\_deF5gs](https://www.youtube.com/watch?v=Bti9_deF5gs) y antes de ingresar a la facultad la próxima clase cómprese un balde extra grande de Pop Corn, se exhibirá en la sala “Los temibles efectos del Data Drifting”

( tiempo estimado 15 minutos, dificultad bajísima)

## Sección Complementaria

### 7. Video : Cambiando la Clase

En el campus clase 3

vea el video

- [Cambiando la clase para mejorar el modelo](#)

( tiempo estimado 7 minutos a 1.5x, dificultad media)

### 8. Corrida Grid Search `clase_binaria1`

Si usted no posee experiencia previa en programación, realice esta tarea junto a un compañero@ que tampoco la posea.

Si ya posee experiencia, trabaje solo.

En el paso 5. usted generó el script [src/rpart/141\\_gridsearch\\_esqueleto.r](#)

Cópielo a [src/rpart/142\\_gridsearch\\_esqueleto\\_binaria2.r](#)

Modifíquelo para que trabaje con `clase_binaria1` en donde `pos = { BAJA+2}` `neg = { BAJA+1, CONTINUA }`

dificultad MUY ALTA

tiempo promedio

- trabajo humano : 30 minutos
- procesamiento : 4 horas

### 9. Análisis de resultados de Grid Search `clase_binaria1`

- Levante el archivo de salida de la corrida anterior en una planilla tipo Excel
- Ordene por `ganancia_promedio` descendente
- El de mayor `ganancia_promedio` decimos que es el primero del ranking
- Cargue los resultados de los rankings 1, 2, 5, 10, 50 y 100 en la hoja GridSearch binaria1 de la Planilla Colaborativa de la Asignatura
- Copie `src/rpart/z101_PrimerModelo.R` a `src/rpart/103_PrimerModelo_binaria1.R` para que trabaje con `clase_binaria1` en donde `pos = { BAJA+2}` `neg = { BAJA+1, CONTINUA }`
- Para cada una de esas seis combinaciones de hiperparámetros, utilizando el script recién modificado genere un submit, súbalo a Kaggle, y copie la ganancia del Public Leaderboard en la correspondiente celda de la planilla
- ¿Cómo compara haber trabajado con la `clase_ternaria` versus `clase_binaria1`?

( tiempo estimado 25 minutos, dificultad alta)

## 10.Zero2Hero segunda parte

Siga en detalle los Jupyter Notebooks carpeta [./src/zero2hero](#) vea los fascículos del zero2hero\_0113 al zero2hero\_0202 inclusive

ejecute los notebooks linea a línea

pregunte a ChatGPT que hace cada línea

experimente cambiar instrucciones y ver que pasa, recuerde siempre hacer una copia del script quitándole la “z” inicial.

Utilice Zulip para preguntar/comentar en el stream [# zero2hero](#)

(tiempo estimado 25 minutos, , dificultad media)

## 11.Videos de Ensembles de Árboles de Decisión

Ver los videos correspondiente a la Clase3 que están en en campus

- Ensembles de Arboles de Decisión
- Algoritmo Random Forest

( tiempo estimado 30 minutos a 1.5x, dificultad media)