**Parte 1**.

Continue considerando regularización de tipo L1 y focalize la atención en el parámero óptimo hallado (C= x). Preste atención al gráfico de importancia de variables (FI) obtenido. Mire cuántas de ellas son variables de conservación genómica y en qué puesto están.

Hint: Son phylop\_100\_vert (puesto 1) y Phascons\_100\_vert (puesto 3).

Son los únicos que aparecen. Quiero entender cuál es el impacto que tiene en el modelo la consideración de features de conservación y entonces decido quitar simultánemamente esas variables.

1. Elimine los feature **phylop\_100\_vert** y  **phastcons100\_vert** del dataset

X,y=data.drop([targetname,'phylop\_phylop100\_vert','phastcons\_phastcons100\_vert'],axis = 1), data[[targetname]]

y vuelva a correr el modelo. Considere el valor de C óptimo y discuta en grupo los siguientes puntos:

* 1. Cayó la performance del modelo?
  2. Las variables que en el modelo original fueron eliminadas por la regularización L1 (coef=0) son o no son importantes? Cuál era el coeficiente de de la variable 'phylop\_phylop100\_vert\_r' antes? Y ahora donde quedó?
  3. Cómo ‘opera’ L1 con las colinealidades del sistema? Qué hace con sus pesos?
  4. Juegue un poco más y elimine también la nueva variable que el modelo considera más importante ahora. De qué tipo es la nueva variable de mayor importancia? Ayúdese con el heatmap del wb0\_exploring.ipynb.
  5. Ejercicio teórico **opcional para el hogar:** considere dos variables perfectamente correlacionadas, vea cuál es la solución óptima de coeficientes que obtiene con regularización L1.

1. Cuantifique ahora correctamente el impacto que tiene la Conservación genómica en el modelo (Hint: revise el wb0\_exploring.ipynb).
   1. Discuta la relevancia de un análisis de clustering preliminar.
2. En base a esta ejercitación discuta: puedo interpretar la importancia de una variable con su coeficiente obtenido en un modelo de regresión logística? Es posible desacoplarla de los restantes coeficientes del sistema?

**Parte 2**

1. Vuelva a considerar todas las variables. Considere la regularización de tipo **L2** y corra todo el modelo.
   1. Discuta qué pasó con la curva de Sparcity vs C. y discútala en términos de lo observado en el gráfico de FI. ¿Qué pasó? ¿Qué variables ‘sobreviven’?
   2. Mirando ahora el gráfico de FI, vea dónde están las variables de conservación genómica. Discuta en su grupo el efecto que tiene la regularización de tipo L2 sobre colinealidades.
   3. Ejercicio teórico **opcional para el hogar:** considere dos variables perfectamente correlacionadas, vea cuál es la solución óptima de coeficientes que obtiene con regularización L2.
   4. Compare el ‘paisaje’ (la curva en este caso univariado) de optimización de hiperparámetros (C) para las regularizaciones L1 y L2. Qué efecto nota en valores de penalización muy elevados?
   5. Ejercicio práctico **opcional para el hogar.** Compare los histogramas de coeficientes en una regularización de tipo L2 para penalizaciones fuertes (C<<1, C=óptimo, y C>>1)