

Instrucciones Notebook (Google Colab)

Este documento recoge información de interés sobre el uso del material contenido en los cuadernos de trabajo (*Notebook*) alojados en el programa / servicio Google Colab.

Qué es Google Colab

Google Colab es una herramienta de Google en la nube para ejecutar código Python o R y crear modelos de Machine Learning. Sería similar a otros recursos más conocidos como Google Docs, donde tenemos un servicio de edición de textos en línea.

La principal ventaja que ofrece esta herramienta es que nos proporciona un entorno de programación totalmente listo para usar desde nuestro navegador web. De este modo, no necesitamos instalar ningún software en nuestro ordenador personal para programar y ejecutar los programas que desarrollemos. Utilizar Google Colab también resulta ventajoso porque libera a nuestro ordenador de tener que ejecutar estos programas, es decir, nos permite ahorrar recursos en nuestro ordenador, y quizás ejecutar códigos para los que nuestro ordenador no tendría recursos suficientemente potentes (CPU, memoria) ejecutar estos programas, es decir, nos permite ahorrar recursos en nuestro ordenador, y quizás ejecutar códigos para los que nuestro ordenador no tendría recursos suficientemente potentes (CPU, memoria). Y todo de forma gratuita.

Para poder hacer uso de este servicio, es obligatorio conectar con una cuenta de Google, bien personal o empresarial. Puede acceder desde cualquier navegador (Chrome, Explorer, Firefox, Safari, etc.), ya que lo único que necesita es tener activa la anteriormente mencionada cuenta de Google.

En resumen: solo necesita un equipo informático básico con una aplicación de navegador Web y una cuenta de Google. No se le pedirá guardar ningún tipo de fichero ni código en su ordenador, por lo que el desarrollo de la actividad será seguro, eficiente y transparente para usted.

MOOC MACHINE LEARNING Y BIG DATA PARA LA BIOINFORMÁTICA

Acceso y uso de Google Colab

Dentro de cada cápsula se encuentra un enlace que le llevará directamente a una página de Google Colab. Recuerde que, para poder hacer uso de este recurso, deberá conectarse a través de una cuenta de Google.

Esta página tendrá un formato similar al de la siguiente imagen:

Modulo3Capsula2_abiertaUGR.ipynb

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda No se pueden guardar cambios

Índice

- Módulo 3 - Ciencia de Datos y Machine Learning.
- ÍNDICE
- 1. ¿QUÉ ES EL APRENDIZAJE SUPERVISADO?
- 2. CARACTERÍSTICAS DE LOS DATOS QUE INFLUYEN EN EL APRENDIZAJE
 - 2.1 Discusión respecto al número de instancias
 - 2.2 Discusión sobre las variables de entrada
 - 2.3 Discusión respecto al sesgo en los datos
- 3. CASO DE ESTUDIO COMPLEMENTARIO: APRENDIZAJE DE CÁNCER DE MAMA MEDIANTE IMÁGENES
- 4. NECESIDAD DE VALIDAR LOS MODELOS DE MACHINE LEARNING
 - 4.1 Casos de estudio en validación
 - 4.2 Tipos de técnicas de validación
 - 4.3 Validación Hold-out
 - 4.4 Validación cruzada de k particiones
 - 4.5 Validación cruzada estratificada

Módulo 3 - Ciencia de Datos y Machine Learning.

3.2 Aprendizaje Supervisado.

Autor:

Por Prof. Alberto Fernández Hilario

Profesor Titular de Universidad de Granada. Instituto Andaluz Interuniversitario en Data Science and Computational Intelligence (DasCI)

Figura 1. Ejemplo de contenido de Notebook en la aplicación online Google Colab

Se pueden observar diferentes bloques en esta página, como son los siguientes:

- Acceso a los menús de programa.

En la parte superior, justo debajo del nombre del documento (acabado en .ipynb) se localizan los menús archivo, editar, insertar, entorno de ejecución, herramientas y ayuda. En principio no es necesario acceder a ninguno de ellos, pero es posible cambiar el estilo de la página pulsando en “Herramientas -> Configuración” para poner el tema “Dark” si resulta más cómodo a la vista.

- Índice.

En la parte izquierda se puede acceder al contenido completo del Cuaderno, pudiendo saltar a las diferentes secciones si es que así se necesita.

- Contenido.

El bloque derecho de la página presenta la información del material de aprendizaje de cada cápsula. Se puede observar que está dividido en una serie de bloques, algunos serán de texto, y otros contendrán código fuente (en Python o R). Su uso se explica con más detalle en la siguiente sección.

Descripción del formato de Notebook

Una vez se acceda a la página / servicio de *Google Colab*, se podrá visualizar el denominado cuaderno Jupyter (*Python Notebook*).

Este cuaderno es un enfoque que combina bloques de texto, junto con bloques o celdas de código. La gran ventaja de este tipo de celdas, es su interactividad, ya que pueden ser ejecutadas para comprobar los resultados directamente sobre las mismas. Por tanto, este material basado en “cuadernos” le guiará paso a paso desde la carga de un conjunto de datos hasta el análisis descriptivo de su contenido.

MUY IMPORTANTE: el orden de las instrucciones (bloques de código) es fundamental, por lo que cada celda de este cuaderno debe ser ejecutada secuencialmente. En caso de omitir alguna, puede que el programa lance un error (se mostrará un bloque salida con un mensaje en inglés de color rojo), así que se deberá comenzar desde el principio en caso de duda. Para hacer este paso más sencillo, se puede acceder al menú “Entorno de Ejecución” y pulsar sobre “Ejecutar anteriores”.

Uso de Notebook en Google Colab

Las celdas de código están resaltadas con un fondo más oscuro y contienen caracteres en una fuente distinta). Para ejecutarlas, se deberá hacer clic en el botón triangular  en la parte izquierda de cada celda de código. Las líneas que comienzan con un “hashtag” (#) son comentarios y no afectan a la ejecución del programa.

Cuando se ejecute el primero de los bloques, aparecerá el siguiente mensaje:

"Advertencia: Este cuaderno no lo ha creado Google.

El creador de este cuaderno es <autor>@go.ugr.es. Puede que solicite acceso a tus datos almacenados en Google o que lea datos y credenciales de otras sesiones. Revisa el código

fuelle antes de ejecutar este cuaderno. Si tienes alguna pregunta, ponte en contacto con el creador de este cuaderno enviando un correo electrónico a <autor>@go.ugr.es.”

No se preocupe, deberá confiar en el contenido del cuaderno (Notebook) y pulsar en "Ejecutar de todos modos". Todo el código se ejecuta en un servidor de cálculo externo y no afectará en absoluto a su equipo informático. No se pedirá ningún tipo de información o credencial, y por tanto podrá seguir con el curso de forma segura.

Para facilitar la ejecución de cada bloque de código, también se puede hacer clic o pinchar sobre cada celda y utilizar el atajo de teclado "ctrl+enter" (cmd+enter en equipos con sistema operativo MacOS).

Al ejecutar el primero de los bloques de código, el sistema conectará con los recursos de Google (entorno de ejecución) y la esquina superior derecha de la ventana de Google Colab pasará de la Figura 2 (No conectado) a la Figura 3 (Conectado). Esto implicará un leve retardo hasta que pueda observar los resultados de la ejecución en su navegador.

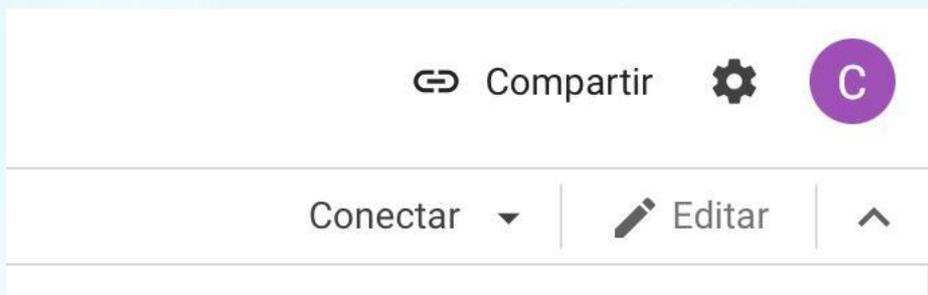


Figura 2. Esquina superior derecha de la ventana de Google Colab, donde se muestra el estado del entorno de ejecución de inicio (No conectado, se muestra la opción "Conectar").

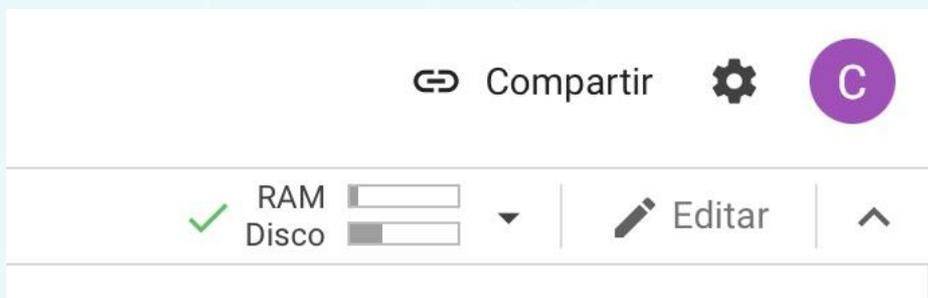


Figura 3. Esquina superior derecha de la ventana de Google Colab, donde se muestra el estado del entorno de ejecución cuando ya está conectado (muestra indicadores de uso de RAM y Disco).

Este entorno estará activo de forma temporal, así que si vuelve a acceder más tarde al Cuaderno, puede que tenga que volver a cargar los recursos del entorno de ejecución y ejecutar bloques de código que ya ejecutó anteriormente para que los bloques siguientes funcionen. En caso de que su entorno de Google Colab se haya desconectado, se mostrará el estado que muestra la Figura 4.

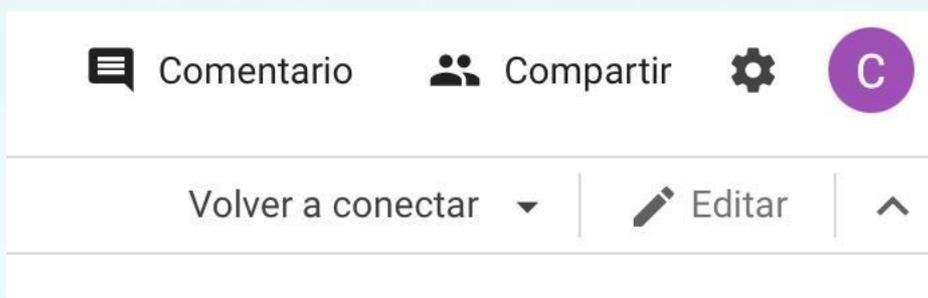


Figura 4. Esquina superior derecha de la ventana de Google Colab, donde se muestra el estado del entorno de ejecución (desconectado).

Cada vez que ejecute un bloque, se podrá observar la salida justo debajo del mismo. La información suele ser siempre la relativa a la última instrucción, junto con todos los `print()` (orden para imprimir) que haya en el código.

En la parte izquierda de una celda ejecutada se puede observar un número entre corchetes. Este número indica el orden en el que se ha ejecutado cada celda. Si se pasa el cursor por esta parte izquierda de la celda también se puede consultar información sobre quién ejecutó la celda, en qué momento y cuánto tardó la ejecución.

En caso de duda o consulta, recuerde que tiene acceso directo al equipo de coordinación del MOOC a través del foro disponible en la plataforma.