[ML on GCP C9] Using pre-trained embeddings with TensorFlow Hub

2 hours Free Rate Lab

Overview

Duration is 1 min

In this lab, you will build a model using pre-trained embeddings from TensorFlow hub.

TensorFlow Hub is a library for the publication, discovery, and consumption of reusable parts of machine learning models. A module is a self-contained piece of a TensorFlow graph, along with its weights and assets, that can be reused across different tasks in a process known as transfer learning, which we covered as part of the course on Image Models.

What you learn

In this lab, you will learn how to:

- How to instantiate a TensorFlow Hub module
- How to find pre-trained TensorFlow Hub modules for a variety of purposes
- How to examine the embeddings of a Hub module
- How one Hub module composes representations of sentences from individual words
- How to assess word embeddings using a semantic similarity test

Setup

For each lab, you get a new GCP project and set of resources for a fixed time at no cost.

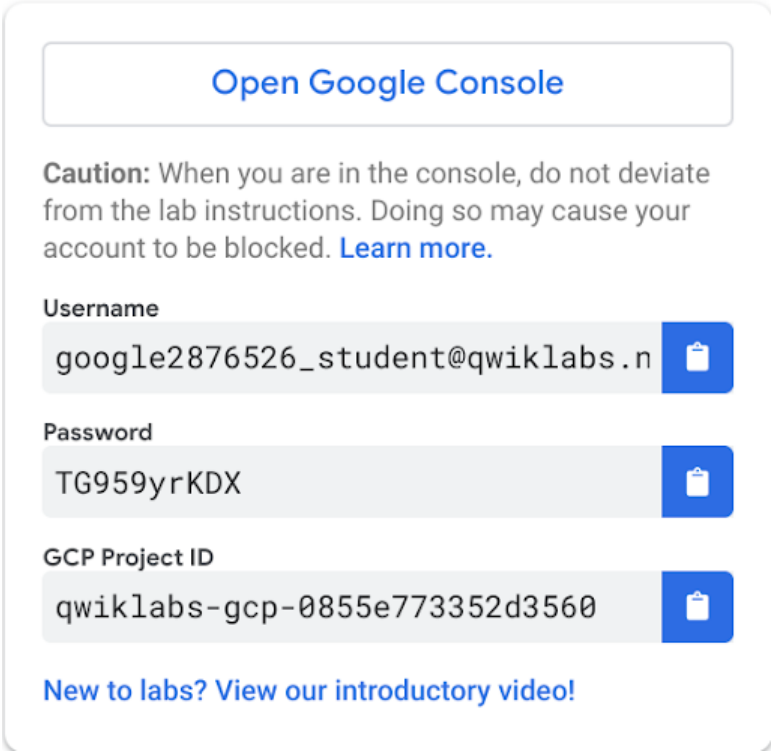
1. Make sure you signed into Qwiklabs using an **incognito window**.

2. Note the lab's access time (for example, 02:00:00 and make sure you can finish in that time block.

   There is no pause feature. You can restart if needed, but you have to start at the beginning.

3. When ready, click START LAB.

4. Note your lab credentials. You will use them to sign in to Cloud Platform Console.



5. Click **Open Google Console**.

6. Click **Use another account** and copy/paste credentials for **this** lab into the prompts.

   If you use other credentials, you'll get errors or **incur charges**.

7. Accept the terms and skip the recovery resource page.

   Do not click **End Lab** unless you are finished with the lab or want to restart it. This clears your work and removes the project.

Create storage bucket

Duration is 2 min

Create a bucket using the GCP console:

Step 1

In your GCP Console, click on the **Navigation menu** (), and

select **Storage**.

Step 2

Click on **Create bucket**.

Step 3

Choose a Regional bucket and set a unique name (use your project ID because it is unique). Then, click **Create**.

Launch Cloud Datalab

To launch Cloud Datalab:

Step 1

Open Cloud Shell. The Cloud Shell icon is at the top right of the Google Cloud Platform web console.

Step 2

In Cloud Shell, type:

```
gcloud compute zones list
```

> **Note**: Please pick a zone in a geographically close region from the following: **us-east1, us-central1, asia-east1, europe-west1**. These are the regions that currently support Cloud ML Engine jobs. Please verify heresince this list may have changed after this lab was last updated. For example, if you are in the US, you may choose **us-east1-c** as your zone.

Step 3

In Cloud Shell, type:

```
datalab create mydatalabvm --
zone <ZONE>
```

Replace with a zone name you picked from the previous step.

Note: follow the prompts during this process.

Datalab will take about 5 minutes to start.

Step 4

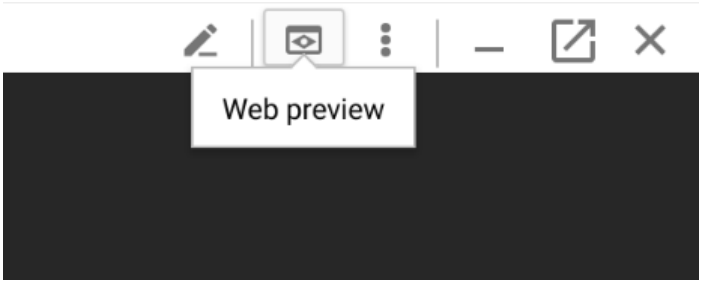
Look back at Cloud Shell and follow any prompts. If asked for an ssh passphrase, hit return (for no passphrase).

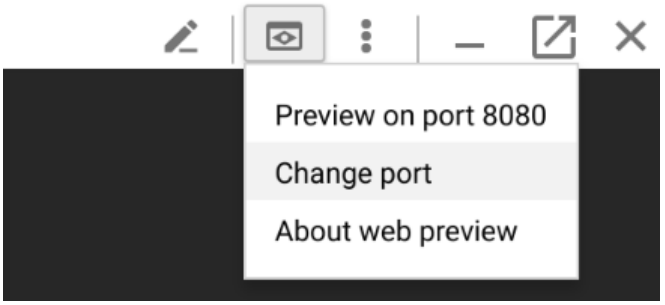
Step 5

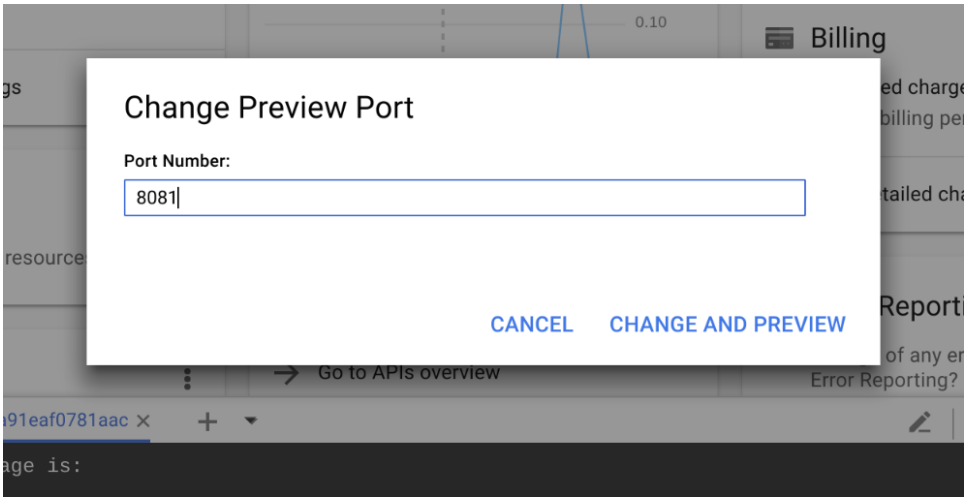
If necessary, wait for Datalab to finishing launching. Datalab is ready when you see a message prompting you to do a **Web Preview**.

Step 6

Click on **Web Preview** icon on the top-right corner of the Cloud Shell ribbon.
Click **Change Port** and enter the port **8081** and click **Change and Preview**.







Note: If the cloud shell used for running the datalab command is closed or interrupted, the connection to your Cloud Datalab VM will terminate. If that happens, you may be able to reconnect using the command **datalab connect mydatalabvm** in your new Cloud Shell.

Clone course repo within your Datalab instance

To clone the course repo in your datalab instance:

Step 1

In Cloud Datalab home page (browser), navigate into **notebooks** and add a new notebook using the icon Notebook on the top left.

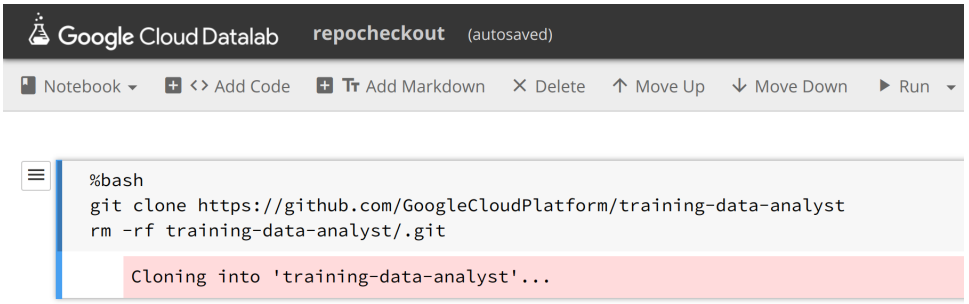
Step 2

Rename this notebook as **repocheckout**.

Step 3

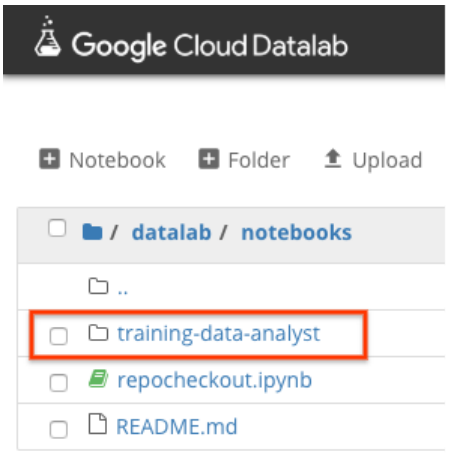
In the new notebook, enter the following commands in the cell, and click on **Run** (on the top navigation bar) to run the commands:

```
%bash
git clone
https://github.com/GoogleCloudPlat
data-analyst
rm -rf training-data-
analyst/.git
```



Step 4

Confirm that you have cloned the repo by going back to Datalab browser, and ensure you see the **training-data-analyst** directory. All the files for all labs throughout this course are available in this directory.



Using pre-trained embeddings with TensorFlow Hub

Duration is 15 min

Step 1

In Cloud Datalab, click on the **Home**icon, and then navigate to **datalab > notebooks > training-data-analyst > courses > machine_learning > deepdive > 09_sequence > labs** and open **reusable-embeddings.ipynb**.

> **Note**: If the cloud shell used for running the datalab command is closed or interrupted, the connection to your Cloud Datalab VM will terminate. If that happens, you may be able to reconnect using the command '**datalab connect mydatalabvm**' in your new Cloud Shell. Once connected, try the above step again.

Step 2

Read through the assignment steps required in the first notebook cell and complete them in your notebook.

If you need more help, you may take a look at the complete solution by navigating to : **datalab > notebooks > training-data-analyst > courses > machine_learning > deepdive > 09_sequence** and open **reusable-embeddings.ipynb**.

End your lab

When you have completed your lab, click **End Lab**. Qwiklabs removes the resources you've used and cleans the account for you.

You will be given an opportunity to rate the lab experience. Select the applicable number of stars, type a comment, and then click **Submit**.

The number of stars indicates the following:

- 1 star = Very dissatisfied
- 2 stars = Dissatisfied
- 3 stars = Neutral
- 4 stars = Satisfied
- 5 stars = Very satisfied

You can close the dialog box if you don't want to provide feedback.

For feedback, suggestions, or corrections, please use the **Support**tab.

Last Tested Date: December 14, 2018

Last Updated Date: December 14, 2018