**Notebooks Jobs de entrenamiento vertex**

Version noviembre 2023

Python 3.10

Todas las pruebas se hicieron corriendo en CPU

No se hicieron pruebas con aceleradores de GPU, parelelismo, hp tunning job ofrecido por vertex

|  |  |  |
| --- | --- | --- |
| Carpetas | Tipo de Job | Detalles adicionales |
| Jobs vertex basic | Script | **Enviar job de entrenamiento – versión más básica en un solo script**:  - crear datos ficticios  - entrenar un modelo  - registrar el modelo en servicio “models” de vertex  - ejecución sin errores |
| Jobs vertex basic v2 | Script | Enviar job de entrenamiento – agregar features adicionales a job anterior. Features nuevas:  - registrar el modelo en servicio “models” de vertex utilizando un **path custom donde se guarda el artefacto del modelo** (guardar el modelo en un path custom y aun así poder registrarlo en vertex)  - **Pasar args al script de entrenamiento**  - **Tener un bucket distinto** entre donde se guarda el script enviado para entrenamiento (archivo .tar.gz en GCS) y un bucket distinto donde se guarda el arterfacto .pkl del modelo  - **Se conecta con otros servicios de GCP para buscar datos** (en el ejemplo GCS pero también se puede BQ)  - Es la versión más completa (comparado con Jobs vertex basic) para enviar Jobs de entrenamiento a GCP |
| Jobs vertex basic v3 | package | Toma lo realizado en el notebook Jobs vertex basic v2 y la única diferencia es que **en vez de pasar un script de entrenamiento, se escribe el código para pasar una serie de scripts de entrenamiento.**  - Tener el código de entrenamiento en **múltiples scripts** los cuales para ser enviados a **vertex primero necesita transformarse en un package .tar.gz y este ser guardado en GCS**  - Generar el package .tar.gz y guardar en GCS **es realizado por el usuario** “manualmente” corriendo el notebook donde se envían los scripts con el código de entrenamiento |
|  |  |  |

**FALTA:**

- enviar Jobs con gpu

- enviar Jobs con código de entrenamiento en paralelo

- enviar job con códigos de entrenamiento que utilizen el servicio de hp tunning de GCP

**DOCUMENTACIÓN INTERESANTE**

Documentación general PYTHON vertex: <https://cloud.google.com/python/docs/reference/aiplatform/latest/google.cloud.aiplatform>

Documentación PYTHON vertex custom training job – script de entrenamiento: <https://cloud.google.com/python/docs/reference/aiplatform/latest/google.cloud.aiplatform.CustomTrainingJob#methods>

Documentación PYTHON vertex custom Python package training job – package de entrenamiento: <https://cloud.google.com/python/docs/reference/aiplatform/latest/google.cloud.aiplatform.CustomPythonPackageTrainingJob>

Documentación github oficial de vertex – notebooks interesantes:

- <https://github.com/GoogleCloudPlatform/vertex-ai-samples/blob/main/notebooks/official/training/get_started_with_vertex_distributed_training.ipynb>

- <https://github.com/GoogleCloudPlatform/vertex-ai-samples/blob/main/notebooks/official/training/hyperparameter_tuning_xgboost.ipynb>

Documentación vertex ai – no tan intensiva en código – sino explicar cosas de vertex:

- variables de ambiente en vertex: <https://cloud.google.com/vertex-ai/docs/training/code-requirements?hl=es-419#environment-variables>

- registrar modelos en menú “models” a partir de un pkl en cierto path de GCS: <https://cloud.google.com/vertex-ai/docs/training/exporting-model-artifacts?hl=es-419#scikit-learn>

- Containers prebuild para entrenamiento: <https://cloud.google.com/vertex-ai/docs/training/pre-built-containers?hl=es-419>

- Containers prebuild para inferencia: <https://cloud.google.com/vertex-ai/docs/predictions/pre-built-containers?hl=es-419>

- VERTEX AI SDK NUEVO: <https://github.com/GoogleCloudPlatform/vertex-ai-samples/tree/main/notebooks/official/vertex_ai_sdk>