Criação e Analise de painéis de IA responsável no AML

Implantar um modelo em um ponto de extremidade em lote

Entender os pontos de extremidade em lote

1. Para obter previsões em lote, você poderá implantar um modelo em um ponto de extremidade
2. Um ponto de extremidade é um ponto de extremidade HTTPS que você pode chamar para disparar um trabalho de pontuação em lote.
3. A vantagem desse ponto de extremidade é que você pode disparar o trabalho de pontuação em lote de outro serviço, como Azure Synapse Analytics ou Azure Databricks
4. Sempre que o ponto de extremidade é invocado, um trabalho de pontuação em lote é enviado para o Workspace do Azure Machine Learning

Para implantar um modelo, primeiro você precisa criar o ponto de extremidade em lote. Para criar um ponto de extremidade em lote, você us uará a classe BatchEndpoint. Os nomes de ponto de extremidade em lote devem ser exclusivos dentro de uma região do Azure.

***from* azure.ai.ml *import* BatchEndpoint**

**#Create a batch endpoint**

**endpoint = BatchEndpoint (**

***name* = "endpoint-exemple"**

***description* = " A batch endpoint"**

**)**

**ml\_client.batch\_endpoints.begin\_create\_or\_update(endpoint)**

Implantar um modelo em um ponto de extremidade em lote

Você pode implantar vários modelos em um ponto de extremidade em lote.

Sempre que você chamar o ponto de extremidade em lote, que dispara um trabalho de pontuação em lote, a implantação padrão será usada, a menos que especificado o contrário.

Executar pontos de extremidade em lote

Usar clusters de computação para implantações em lote

Para processar os novos dados em lote paralelos, será necessário provisionar um cluster de cálculo com mais de uma instância máxima.

Para criar um cluster de computação, você poderá usar a classe AMLCompute

***from* azure.ai.ml.entities *import* AmlCompute**

**cpu\_cluster = AmlCompute (**

***name* = "aml-cluster",**

***type* = "aml-compute",**

***size* = "STANDARD\_DS11\_V2",**

***min\_instances* = 0,**

***max\_intances* = 4,**

***idle\_time\_before\_scale\_down*=120,**

***tier* ="Dedicate",**

**)**

**cpu\_cluster =**

**ml\_client.compute.begin\_create\_or\_update(cpu\_cluster)**

Implantar seu modelo do Mlflow em um ponto de extremidade em lote

Implantar um modelo do Mlflow em um ponto de extremidade

* Para implantar um modelo do Mlflow em um ponto de extremidade em lote, você usurá a classe BarchDeployment.
* Quando você implanta um modelo, precisa especificar o comportamento do trabalho de pontuação em lote.
* Ao configurar a implantação do modelo, você poderá especificar:
  + instance\_count
  + max\_concurrency\_per\_instance
  + mini\_batch\_size
  + output\_action
  + output\_file\_name

Implantar um modelo personalizado em um ponto de extremidade em lote

Se você quisaer implantar um modelo em um ponto de extremidade em lote, sem usar o formato de modelo do Mlflow, precisará criar o script de pontuação e o ambiente.

1. Criar o script de pontuação: O script de ponutação é um arquivo que lê os novos dados, carrega o modelo e executa a pontuação.
2. Criar um ambiente: Sua implantação requer um ambiente de execução no qual executar o script de pontuação. Qualquer dependência necessária pelo código deve ser incluída no ambiente.
3. Configurar e criar a implantação: Finalmente, você pode configurar e criar a implantação com a classe BatchDeployment.

Invocar pontos de extremidade em lote

* Para preparar dados para previsões em lote, você poderá registrar um pasta com um ativo de dados no workspace do Azure Machine Learning.
* Você poderá usar o ativo de dados registrado como entrada ao invocar o ponto de extremidade em lote com o SDK do Python

***from* azure.ai.ml *import* Input**

***from* azure.ai.ml.constants *import* AssetType**

**input = Input(*type* = AssetType.URI\_FOLDER,**

***path*="azureml:new-data:1")**

**job = ml\_client.batch\_endpoints.invoke(**

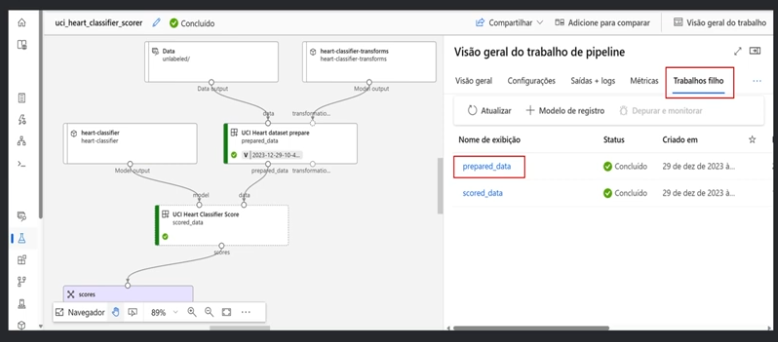
***endpoint\_name* = endpoint.name,**

***input* = input**

**)**

Resolver um problema de um trabalho de pontuação em lote

O trabalho de pontuação em lote é executado como um trabalho de pipeline. Se você quiser solucionar problemas do trabalho de pipeline, poderá examinar seus detalhes e as saídas e logs do próprio trabalho de pipeline.



Revisão do modulo

Saiba como implantar um modelo em um ponto de extremidade online gerenciado para previsões em tempo real.

Como implantar um modelo em um ponto de extremidade em lote para previsões em lote.

Materiais de Apoio

Os materiais complementares e de apoio que oferecemos têm como objetivo fornecer informações para facilitar e enriquecer a sua jornada de aprendizado no curso "Implantação de Modelos em Pontos de Extremidade em Lote". Aqui você encontrará links úteis, como slides, repositórios e páginas oficiais, além de dicas sobre como se destacar na DIO e no mercado de trabalho 😉

Recursos Adicionais

Durante este conteúdo, compreendemos os fundamentos da engenharia de prompts. Para ajudá-lo a aprofundar o conhecimento, disponibilizamos a seguir o material complementar contendo os conteúdos e links apresentados no curso:

Slide: Módulo 6 - Implantar e Consumir Modelos com o Azure Machine Learning.pptx

Dicas e Links Úteis

Para se desenvolver ainda mais e se destacar na DIO e no mercado de trabalho, sugerimos os seguintes recursos:

Artigos e Fórum da DIO: Compartilhe seus conhecimentos e dúvidas através dos artigos (visíveis globalmente na plataforma da DIO) e nos fóruns específicos para cada experiência educacional, como nossos Bootcamps.

Rooms: Participe do Rooms, uma ferramenta de bate-papo em tempo real onde você pode interagir com outros participantes dos nossos Bootcamps, compartilhando dúvidas, dicas e snippets de código.

Exploração na Web: Utilize motores de busca para aprofundar seu conhecimento sobre temas específicos. Páginas como o StackOverflow são recursos valiosos para encontrar soluções e expandir seu entendimento.

Com esses materiais complementares, você estará bem equipado para explorar todo o potencial e se destacar em suas iniciativas. Continue aproveitando as oportunidades de aprendizado, e não hesite em buscar mais conhecimento e compartilhar suas descobertas com a comunidade!

