

Regression Tree

Michele Andreucci

Settembre 2020

1 Introduzione

Nel data mining la regressione consiste nel predire un valore di un attributo numerico associato a un esempio sulla base di valori osservati per altri attributi dell'esempio medesimo. Un albero di regressione viene costruito utilizzando tecniche di apprendimento a partire dall'insieme dei dati iniziali (training set) per i quali è nota la classe.

2 RT-Miner

RT-Miner è composto da due applicazioni, un client ed un server. Lo scopo del progetto è permettere a molteplici client di eseguire un algoritmo di RT su un server utilizzando come sorgente dei dati un database MySQL.

2.1 Server

Il server si occupa principalmente di accettare le richieste dei client. Esso esegue l'algoritmo di RT vero e proprio utilizzando come sorgente dei dati una tabella (specificata dal client) di un database MySQL.

Le funzionalità del server sono le seguenti:

- Inizializza il server su una porta arbitraria.
- Capacità di gestire le richieste di molteplici client contemporanea.
- Leggere il contenuto di una tabella di una base di dati indipendentemente dal numero di attributi.
- Eseguire l'algoritmo di Regression Tree Miner.
- Salvare su file l'esito della computazione.
- Restituire ai client l'esito della computazione

2.2 Client

Il client permette ad un'utente di connettersi al server e selezionare la fonte dei dati su cui eseguire l'algoritmo di RT.

2.3 Estensione

L'estensione principale rispetto al progetto originale è la grafica per l'applicazione client.

2.4 Grafica

La grafica per l'applicazione client è stata scritta utilizzando le JavaFX con l'ausilio dei FXML, che aprendoli con SceneBuilder, uno strumento di layout visivo, ti consentono di progettare UI dell'applicazione JavaFX.

L'interfaccia grafica permette all'utente di connettersi al server ed eseguire operazioni di mining.

3 Manuale utente

3.1 Server

Per avviare il server è necessario eseguire il file batch(.bat) presente nella cartella /Progetto/distribuzione/server.bat. Di default, la porta utilizzata dal server è la 8080. Tuttavia, al momento dell'esecuzione è possibile specificare la porta da utilizzare da riga di comando come parametro input al file (.Bat).

Verrà stampato un errore se il formato della porta non è valido oppure se la porta è occupata.

3.2 Client

Per avviare il client è necessario che il server sia attivo.

Va quindi eseguito il file .bat presente nella cartella /Progetto/distribuzione/progettoBase/client.bat.

3.3 Esempio di uso Client

1. se la connessione va a buon fine visualizzeremo una schermata come questa

```
Socket[addr=/127.0.0.1,port=8080,localport=56457]
Learn Regression Tree from data [1]
Load Regression Tree from archive [2]
```

2. selezioniamo una operazione e inseriamo il nome tabella.
3. Ci verrà richiesto se vogliamo ripetere un'operazione sulla stessa tabella.
4. In caso negativo potremo decidere se selezionare una nuova operazione nel menu.

MainTest [Java Application] C:\Program Files\Java\jre

```
Learn Regression Tree from data [1]
Load Regression Tree from archive [2]
```

1

Table name:

provaC

Starting data acquisition phase!

Starting learning phase!

Starting prediction phase!

0:X=A

1:X=B

0

0:Y<=2.0

1:Y>2.0

1

Predicted class:1.5

Would you repeat ? (y/n)

y

Starting prediction phase!

0:X=A

1:X=B

0

0:Y<=2.0

1:Y>2.0

0

Predicted class:1.0

Would you repeat ? (y/n)

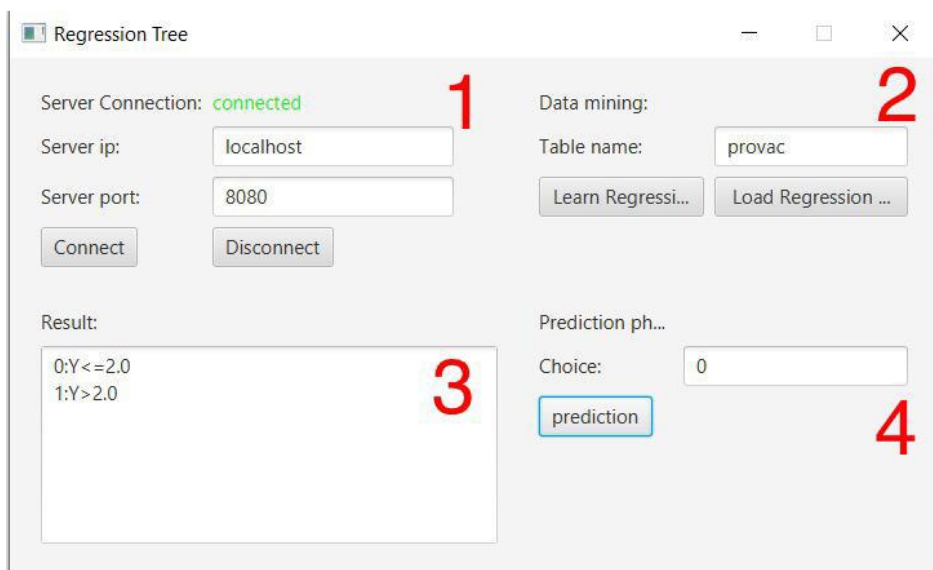
3.4 ClientJavaFX

Per avviare il client interfaccia grafica, è necessario che il server sia attivo. Va quindi eseguito il file batch presente nella cartella /Progetto/distribuzione/progettoEsteso/client.bat.

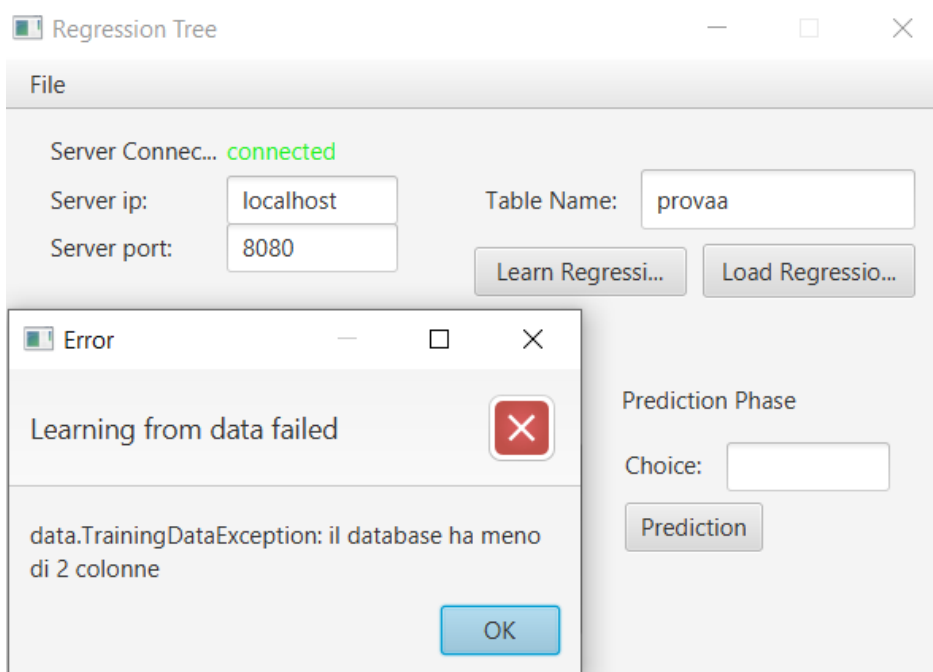
1. L'utente sceglie IP e porta del server a cui desidera connettersi. Un'etichetta posta in cima al form comunica lo stato della connessione. Se IP o porta risultano non validi si aprirà una finestra di errore. L'utente può

scegliere, attraverso i pulsanti, di connettersi o disconnettersi. Quando si è disconnessi e si prova a fare un'operazione verrà mostrato a video un errore.

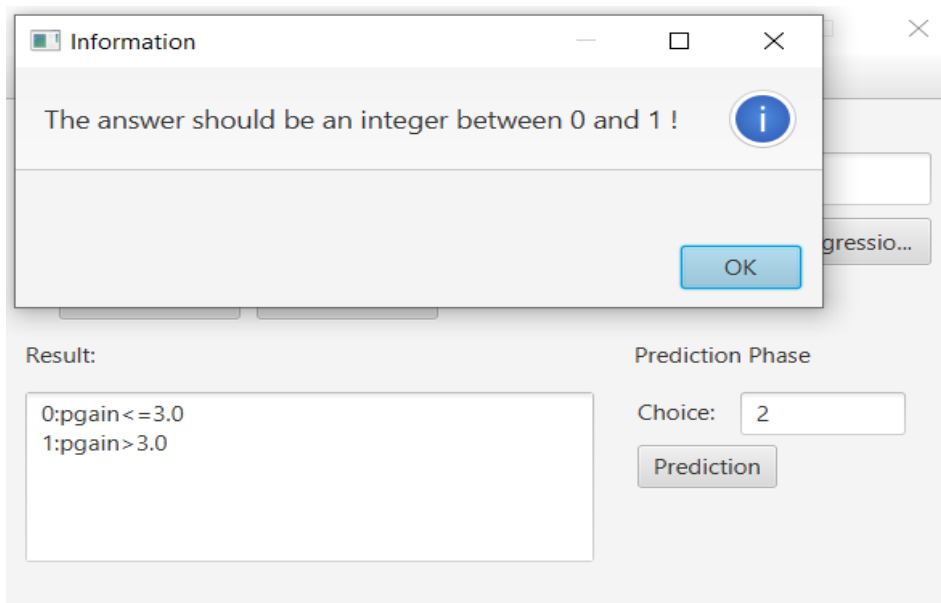
2. L'utente sceglie nome della tabella del database da utilizzare. Se la tabella risulta non presente nel database viene mostrato a video un errore. L'utente può decidere se eseguire l'algoritmo di mining partendo da una tabella del database oppure leggere da file.
3. Il risultato, in forma testuale, dell'operazione di data mining.
4. L'utente può effettuare la fase di predizione.
5. Ottenuta la predizione, l'utente può utilizzare il tasto Load per ripetere il processo di esplorazione del Regression Tree



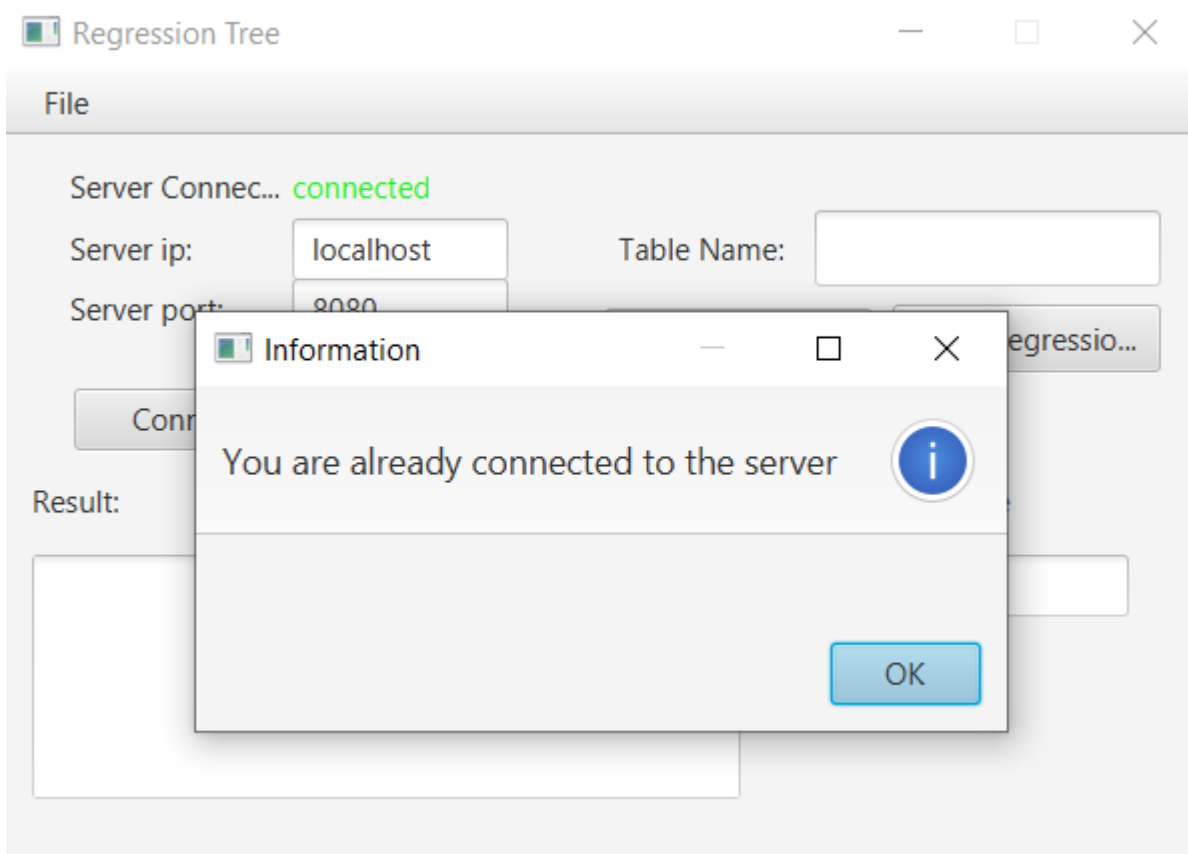
Se l'utente inserisce il nome di una tabella inesistente, compare una schermata di errore



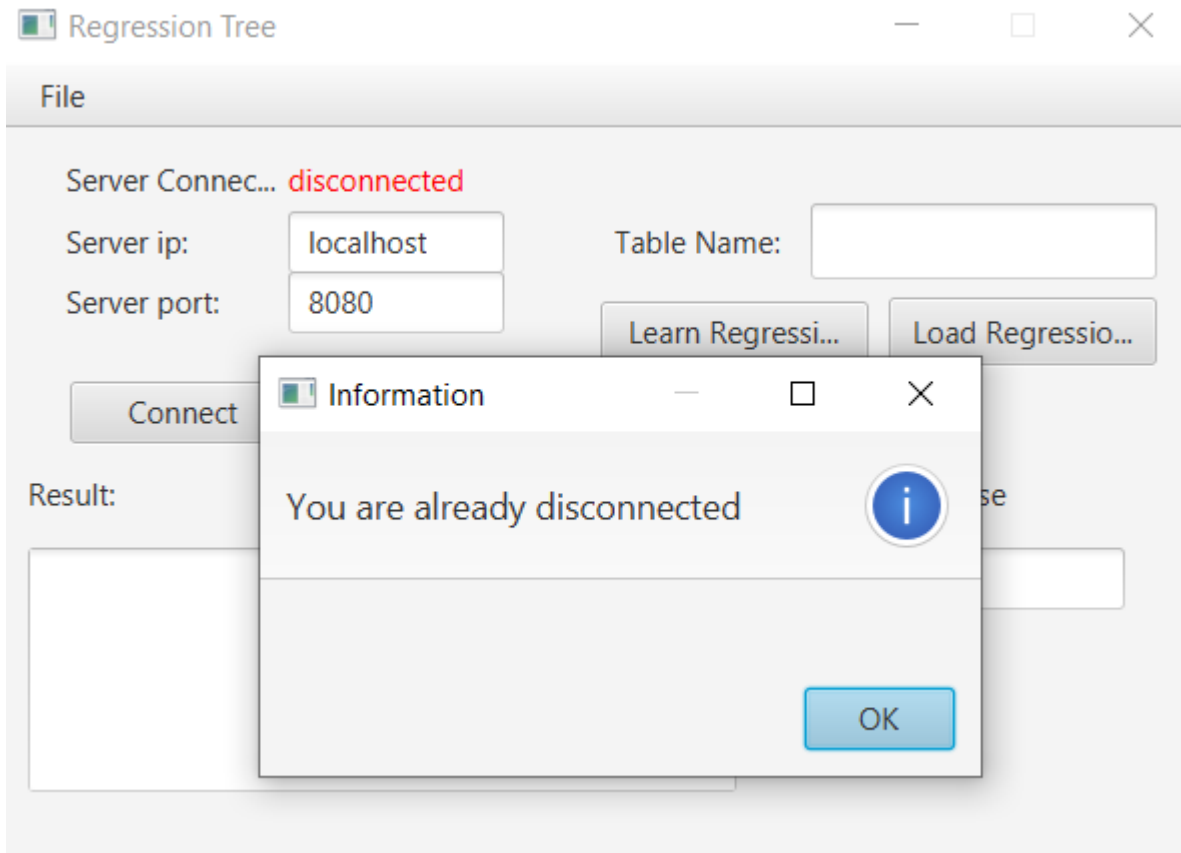
Se l'utente effettua una scelta sbagliata, compare una schermata d'avviso



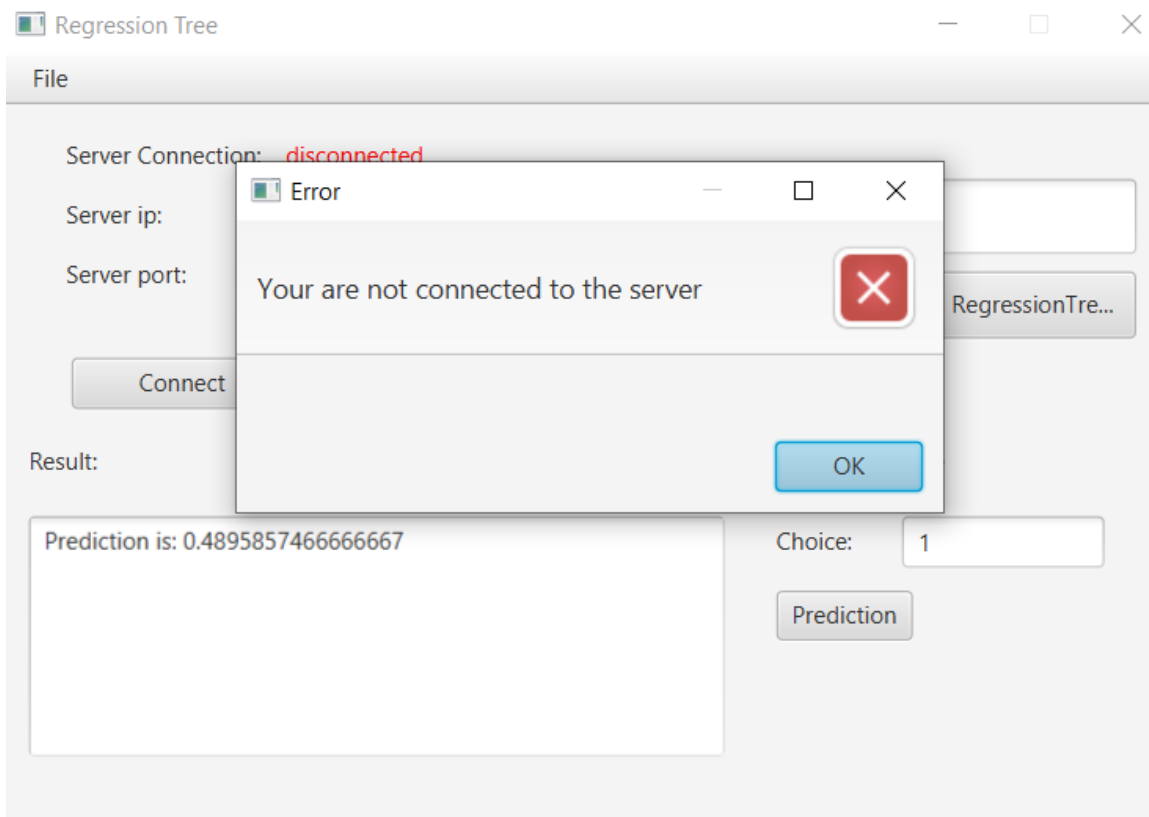
Se l'utente cerca di connettersi al server quando è già connesso esce una schermata di avviso



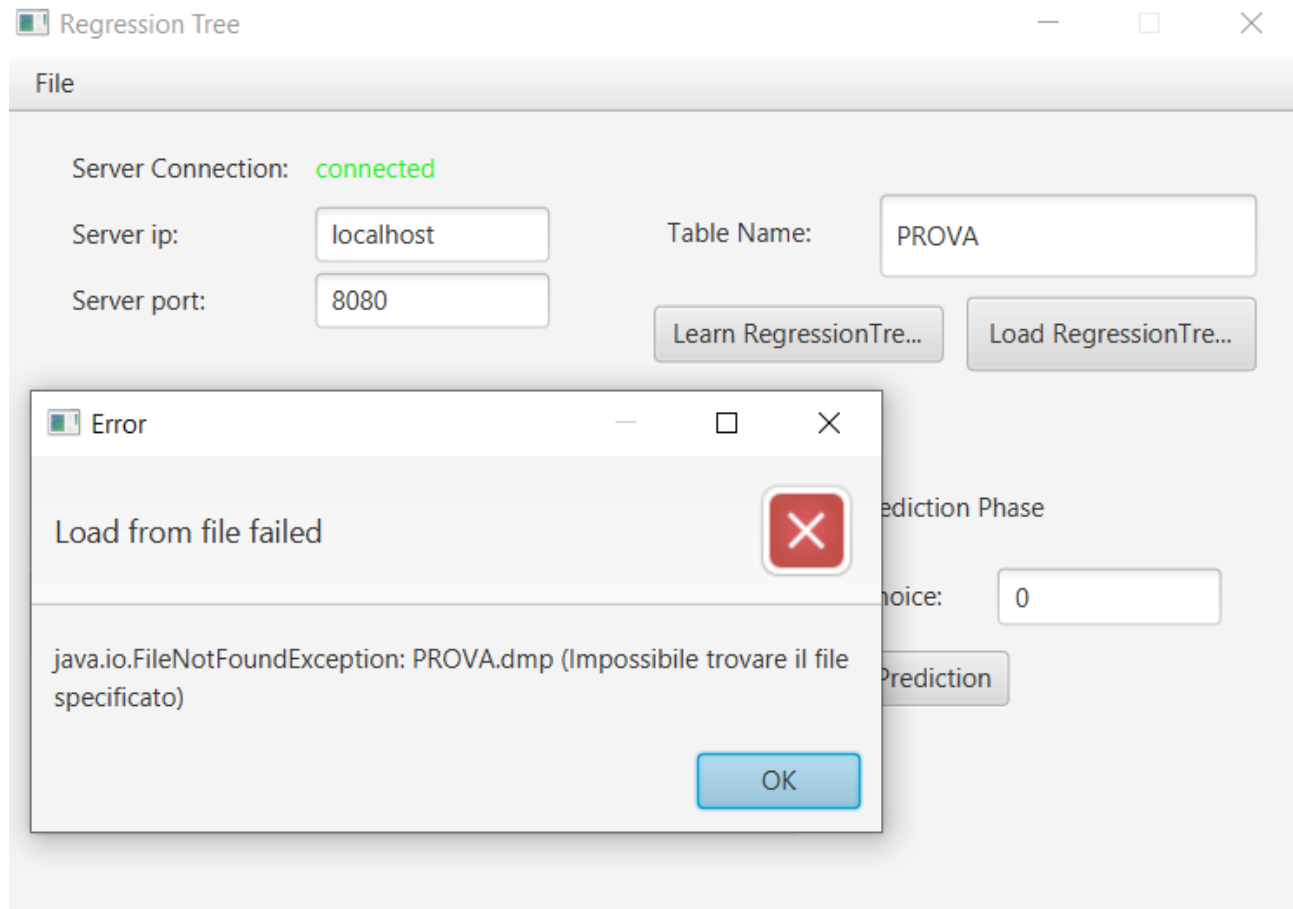
Se l'utente cerca di disconnettersi dal server quando è già disconnesso esce una schermata d'avviso



Se l'utente cerca di salvare o caricare l'albero di regression dal database mentre non è connesso al server compare una schermata di errore



Se l'utente cerca di caricare un file che non esiste compare una schermata di errore



4.1 Guida per l'installazione

Installare JRE 8. I file per la distribuzione (file .Bat e .jar) sia per il client che per il server si trovano in /Progetto/distribuzione/

Per eseguire il server è necessario aver installato MySQL Command Line 5.1 sul proprio sistema operativo. Dopodichè bisogna accertarsi che il servizio di MySQL sia in esecuzione. Bisogna poi inizializzare il database MySQL eseguendo lo script mapdb.sql presente nella cartella /Progetto/distribuzione/

Lo script SQL contiene le istruzioni per creare un database MapDB contenente una tabella di esempio chiamata provaC. Inoltre esso crea un nuovo utente (se non presente) chiamato MapUser con diritti di accesso per il database MapDB.