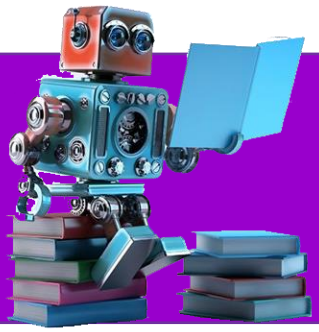


عنوان مقاله: مقدمه‌ای بر سرویس‌های یادگیری ماشین در SQL Server 2017
نویسنده مقاله: مجید جوادی
تاریخ انتشار: اسفند ۱۳۹۶
منبع: http://nikamooz.com/machine-learning/

مقدمه‌ای بر سرویس‌های یادگیری ماشین در SQL Server 2017

در سال‌های اخیر شاید واژه یادگیری ماشین یا داده کاوی و یا حتی تحلیل داده را شنیده باشید. اما به طور مشخص یادگیری ماشین و دلایل همه جا گیر شدن آن چیست؟ اجازه دهید برای درک بهتر این بحث از هدف شکل‌گیری آن استفاده کنیم. هدف یادگیری ماشین شناخت ساختار داده است، به گونه‌ای که بتوان با شناخت مناسب، مدلی ایجاد کرد که درک هرچه بهتر آن مدل به تصمیم گیری ما در فرایند کسب و کار و موفقیت کمک کند.

مایکروسافت به منظور توسعه اپلیکیشین‌های هوش مصنوعی چارچوبی با عنوان سرویس یادگیری ماشین ارائه داده است. دنیاي فناوري اطلاعات سالهاست که شاهد معرفي زبان‌هاي برنامه‌نويسي به منظور انجام فعاليت‌ها و اهداف مختلف مي‌با شد. يكي از زبان‌هاي خاص منظوره در حوزه داده كاوی، یادگیری ماشین و تحلیل آماري دادگان كه در طی چند سال گذشته بسیار مورد توجه قرار گرفته است، زبان R مي‌با شد. از امكانات فراهم شده توسط شركت مایکروسافت در این خصوص، بکارگیری زبان قدرتمند و غنی این زبان در محیط پایگاه داده معروف و محبوب SQL Server می‌باشد. R زبان برنامه‌نويسي آماری می‌باشد که غالبا به منظور تحلیل و محاسبات آماری، دادهکاوی و یادگیری ماشین از آن استفاده می‌شود. علاوه بر حوزه‌های کاری چشمگیر یاد شده، ابزاری بسیار قدرتمند در بصری‌سازی داده و اشکال گرافیکی به حساب می‌آید.

پیش از معرفی بیشتر ابزار یادگیری ماشین در SQL Server ابتدا برخی از ویژگی‌های بارز آن معرفی می‌شود. یکی از قابلیت‌های اصلی ابزار یادگیری ماشین به کارگیری پکیج‌های ارائه شده برای این منظور است. از این رو مدیریت پکیج‌ها به طور قابل توجهی در SQL Server بهبود یافته است. این پکیج‌ها با جامعه بیش از هزاران پکیج سورس باز فعال هستند. در نتیجه نصب و پاک کردن این بسته lll و نهایتا کنترل آنها بسیار راحت می‌باشد. قابلیت پشتیبانی از چند سکویی بودن نیز از دیگر ویژگی با اهمیت ابزار یادگیری ماشین در SQL Server می‌باشد. پیش از تجمیع R و SQL Server، بسیاری از کاربران و نیز سازمان‌ها مشکلات زیادی برای رسیدن به موفقیت در تحلیل داده داشتند. زیرساخت‌های SQL Server databases, roles, access, security موجود کمک می‌کنند تا در محیط‌های تجاری با حجم انبوهی از اطلاعات، داده‌کاوی به خوبی انجام شود.

از دیگر قابلیت های کلیدی یادگیری ماشین فراهم شده توسط SQL Server می‌توان به موارد زیر اشاره کرد:

- تحلیل داده بسیار منعطف‌تر، به ویژه در انباره داده حجیم.
- تحلیل دیتاست‌های بسیار بزرگ
- تحلیل هر نوع کلان داده
- به اشتراک‌گذاری بسیار سادتر داده
- غلبه بر محدودیت حافظه
- موازی‌سازی هرچه بهتر اجراء

معرفی اجزای کلیدی یادگیری ماشین

مایکروسافت به منظور بهره گیری هر چه بهتر از زبان R، استفاده از آن را به طور کلی در دو نسخه عمومی(Community) و دیگری نسخه تجاری(Enterprise) تقسیم‌بندی کرده است. همانگونه که در تصویر قابل مشاهده است هر یک با ارائه راهکار مناسب، استفاده متناسبی را عرضه می‌کنند که در ادامه با بررسی بیشتر، هر یک را شرح می‌دهیم.

خانواده محصولات زبان R که توسط شرکت مایکروسافت ارائه شده است به شرح ذیل می‌باشد:

- Microsoft R Open
- Microsoft R Client
- Microsoft R Server
- SQL Server R Services

Microsoft R Open

این نسخه کاملا متن باز بوده و توسط شرکت مایکروسافت عرضه شده است تا وظایف تحلیل آماری و علوم داده را انجام دهد. علاوه بر رایگان بودن این نسخه می‌توان به قدرت بالای سازگاری آن با دیگر موتورهای R نظیر R studio اشاره کرد. نکته قابل توجه این است که با استفاده از کتابخانه MKL (Math Kernel Library) برای انجام عملیات محاسباتی برداری و ماتریسی، امکان استفاده از پردازش‌های چند نخی با کارایی بالا را فراهم می‌آورد. متاسفانه این نسخه مشکل محدودیت حافظه را دارد، یعنی تنها داده‌های که درون حافظه وجود دارند را پردازش می‌کند و این انتقال داده به حافظه بسته به کامپیوتر مجری دارد. لازم به ذکر است که R Open روی تمام نسخه‌های SQL Server به استثنای نسخه Express اجرا خواهد شد.

Microsoft R Client

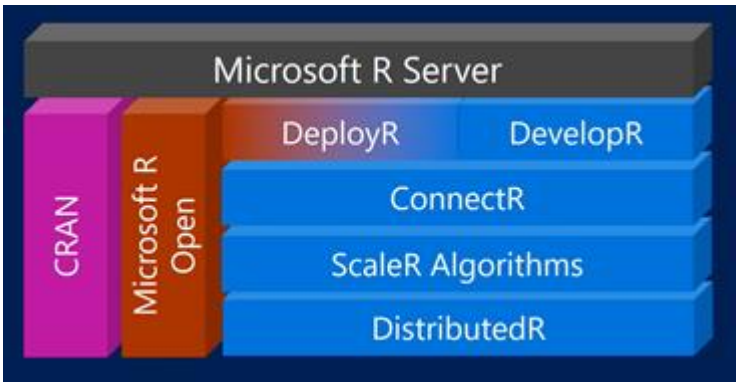
این نسخه از نرم‌افزار تقریبا مشابه نسخه R Open بوده با این تفاوت که برای پردازش‌های موازی سنگین و محاسبات چند نخی کتابخانه RevoScaleR را معرفی کرده است. این کتابخانه از تکنولوژی ScaleR و خصوصیات آن برای محاسبات موازی استفاده می‌کند. این نسخه محدودیت حافظه‌ی محلی را دارد. با وجود اینکه توابع ScaleR می‌توانند از محاسبات موازی استفاده کنند، اما صرفا نظر از داشتن بیش از دو هسته پرداز شی، پردازش تنها به دو رشته کنترلی(نخ) محدود می‌شود. و محدودیت دیگر اینکه تمام محاسبات به قابلیت‌های کامپیوتر کلاینت نظیر دیسک، حافظه محدود می‌شود.

Microsoft R Server

تقریبا پراستفاده‌ترین نسخه از خانواده محصولات تحلیل داده مایکروسافت R Server می‌باشد. به عنوان اولین نکته باید توجه داشت که بیشتر قابلیت‌های نسخه‌های R Open و R Client در این نسخه نیز وجود دارد و از این نسخه به طور ویژه در اهداف تجاری استفاده می شود. مشابه نسخه‌های R Open و R Client، این نسخه نیز تمامی فعالیت‌های داده‌کاوی، تحلیل داده و یادگیری ماشین را انجام می‌دهد. این نسخه در مقابل نسخه‌های یاد شده محدودیت حافظه را نداشته و قابلیت مقیاس‌پذیری دیسک نیز در این نسخه وجود دارد. باید توجه کرد که از

[۱] open source
[۲] multi-threaded

R Server و R Client تنها در نسخه‌های Enterprise یا Developer مربوط به SQL Server می‌توان استفاده کرد. از این نسخه زمانی استفاده می‌شود که نخواهیم دستورات R را مستقیا درون T-SQL اجرا کنیم.



شکل فوق اجزای تشکیل دهنده Microsoft R Server را نمایش می‌دهد که در ادامه برخی از ویژگی‌های مهم آن توضیح داده می‌شود:

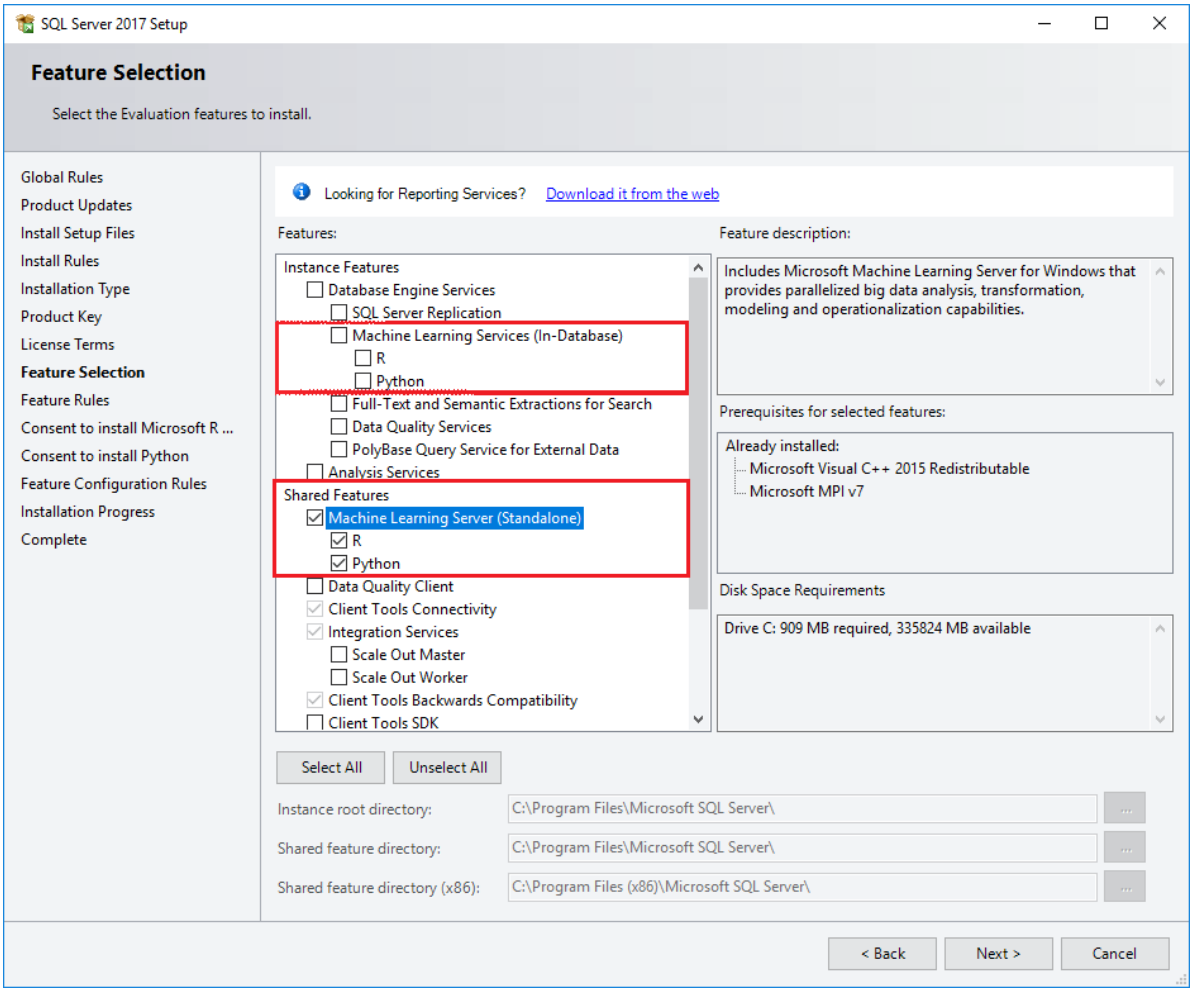
- DeployR یک تکنولوژی تجمیعی برای استقراء تحلیل‌های R درون وب، دسکتاب، موبایل و داشبورد مدیریتی و نیز سیستم‌های تحت سرور چند سکویی می‌باشد.
- ConnectR با این مولفه این قابلیت فراهم می‌شود که اتصال به هر منبع داده‌ای با سرعت بسیار بالا انجام شود.
- DistributedR به منظور سازگاری در اجرای موازی این چارچوب ارائه شده ا ست که شامل سرویس‌های برای ارتباطات، ادغام منابع ذخیره‌سازی و مدیریت حافظه می‌باشد.
- ScaleR این مولفه ارائه دهنده الگوریتم‌های بهینه شده برای اجرای موازی کلان داده می‌باشد. از قابلیت‌های کلیدی این مولفه می‌توان به حذف محدودیت حافظه و پنهان‌سازی اجرای توزیعی از دید کاربر یاد کرد.

SQL Server R Services (in-database Microsoft R Server)

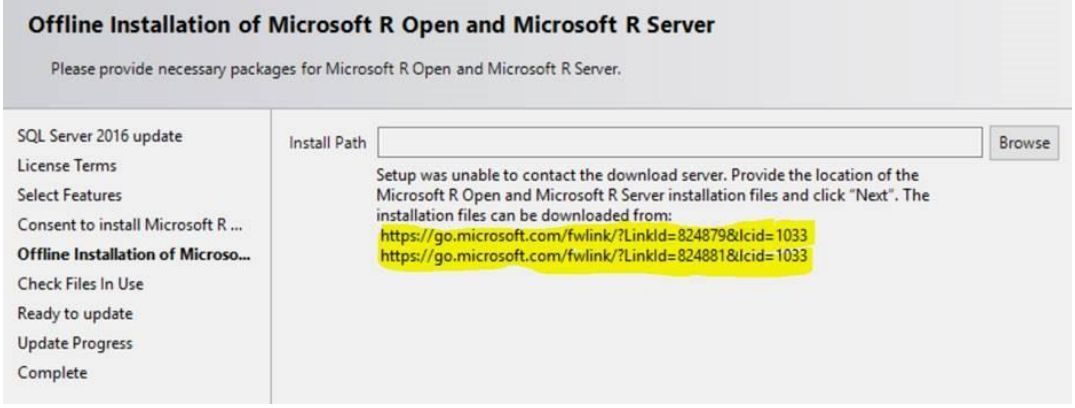
این نسخه تقریبا نسخه R Server است که درون موتور دیتابیس SQL Server قرار گرفته و اجرای الگوریتم‌های ScaleR برای محیط‌های با کارای بالا و مقیاس پذیر از ویژگی بارز این نسخه به حساب می‌آید. مدیریت حافظه و دیسک به عهده‌ی خود SQL Server می‌باشد. به منظور پشتیبانی زبان R درون محیط SQL Serve، سرویسی به نام SQL Server Trusted Launchpad به خود SQL Server اضافه می‌شود.

نکات لازم در مراحل نصب

اولین گام ضروری در هنگام نصب، انتخاب نوع تعامل R و SQL Server می باشد. دو انتخاب در هنگام نصب وجود دارد. یکی نصب سرور یادگیری ماشین به طور مستقل (Standalone)و دیگری درون-دیتابیسی-(In Database) که در این روش موتور R درون خود Database تعبیه شده است. زمانی که Standalone انتخاب شود، برای اجرای R باید از Microsoft R Server استفاده کرد. در صورت انتخاب Machine learning Server(In-Database) باشد باید از قابلیت‌های نسخه SQL Server R Services استفاده کرد.



در صورتی که امکان دانلود نسخه‌های R Server و R Open فراهم نباشد، این صفحه ظاهر شده و انتخاب گزینه‌ی مورد نظر برای نصب به صورت آفلاین را فراهم می‌آورد. باید توجه کرد که آدرس دانلود فایل‌های مورد نیاز در قسمت مشخص با رنگ زرد در تصویر نمایش داده شده است.



در صورتی که تمام مراحل به موفقیت انجام شده باشد، نصب به پایان رسیده و آماده بهره برداری از آن فراهم می‌شود.

بکارگیری R در SQL Server

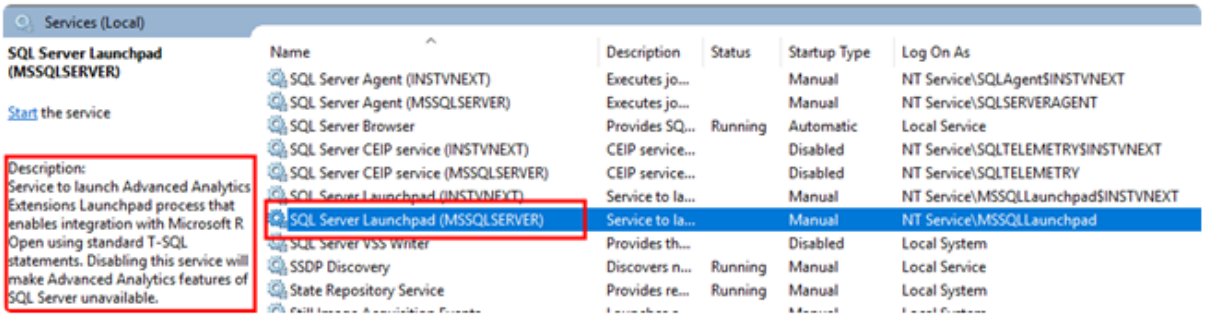
در این مرحله هدف صرفا بررسی اجمالی اجرای دستورات در SQL Server است. به منظور درک بهتر مثال ساده‌ای در ادامه دنبال می‌شود که تحلیل داده‌ای روی داده‌های عددی انجام شده و نتایج را به SQL Server باز خواهیم گرداند.

پس از اینکه نصب با موفقیت انجام شد می‌بایست پیکربندی لازم را برای اجرای دستورات R انجام شود. با اجرای دستور زیر فعال‌سازی انجام می‌شود. باید توجه داشت که این دستور تنها توسط کاربران با مجوز Admin روی سرور قابل اجرا است.

```
EXEC SP_CONFIGURE 'external scripts enabled', 1;
GO

RECONFIGURE;
GO
```

بعد از اجرای دستور فوق نیاز به Restart کردن نمونه SQL Server می‌باشد. پس از انجام این اقدام مطمئن شوید که سرویس SQL Server Launchpad حتما اجرا می‌باشد.



برای اجرای دستورات R درون T-Sql قالب مشخصی باید استفاده شود. در ابتدا رویه Sp_Execute_External_Script را اجراء می کنیم و با مشخص کردن ورودی‌های لازم، دستور اجرا کامل می شود. وردی @language مشخص می‌کند که برای تحلیل داده قرار است از زبان تحلیلی R استفاده کرد یا از عمومی Python. از آنجایی که هدف این مقاله اجرای دستورات R می‌باشد، زبان مورد نظر را با R مشخص می‌کنیم. ورودی بعدی Script است که تمامی دستورات لازمR در این قسمت قرار می‌گیرند. در این مثال ساده قرار است دنباله‌ی از اعداد بین ۱ تا ٤ با فاصله‌ی افزایشی ۰٫۵ ایجاد شود.

```
EXEC sp_execute_external_script
        @language = N'R',
        @script = N'OutputDataSet <- data.frame(seq(1,4,0.5));';
GO
```

پس از اجرای دستور فوق، خروجی به شکل زیر نمایش داده می‌شود.

(No column name)
1
1.5
2
2.5
3

3.5
4

اجازه دهید کمی بیشـــتر با این قابلیت آشـــنا شـــویم. یکی از قابلیت‌های که یادگیری ماشـــین SQL Server ارائه می‌دهد، تعامل آن با R در راستای اجرای دستورات بر روی داده‌های درون SQL Server می‌باشد. در این مثال برای اجرای دستورات R اطلاعاتی از جدول فروش شخصی پایگاه داده Adventuerwork2016 اطلاعاتی را واکشی کرده و نمایش می‌دهد.

```
DECLARE @rscript NVARCHAR(MAX);
SET @rscript = N'OutputDataSet <- InputDataSet;';
DECLARE @sqlscript NVARCHAR(MAX);
SET @sqlscript = N'
SELECT FirstName, LastName, SalesYTD
FROM Sales.vSalesPerson
WHERE SalesYTD >= 2000000
ORDER BY SalesYTD DESC;';
EXEC sp_execute_external_script
@language = N'R',
@script = @rscript,
@input_data_1 = @sqlscript;
GO
```

همانطور که در تصـــویر بالا مشـــخص اســـت دو متغییر Rscript و SqlScript را تعریف کرده‌ایم که در Rscript د ستورات مربوط به R و درون متغیر SqlScript د ستورات مربوط به واکشی اطلاعات از SQL Server قرار خواهد گرفت. سپس با اجرای رویه Sp_Execute_External_Script دستورات لازم اجراء می‌شود. با اجرای دستورات فوق خروجی به شکل زیر قابل مشاهده خواهد بود.

(No column name)	(No column name)	(No column name)
Linda	Mitchell	4251368.5497
Jae	Pak	4116871.2277
Michael	Blythe	3763178.1787
Jillian	Carson	3189418.3662
Ranjit	Varkey Chudukatil	3121616.3202
José	Saraiva	2604540.7172
Shu	Ito	2458535.6169
Tsvi	Reiter	2315185.611

منبع

۱- کتاب داده‌کاوی کاربردی با R – نویسندگان (مجید جوادی، محمد مرادی، سهیلا مهر مولایی)

سرویس های یادگیری ماشین درsql server ،آموزشsql ، آموزشsql server
sql server،معرفی اجزای کلیدی یادگیری ماشین، آشنایی با اجزای کلیدی یادگیری ماشین،آشنایی با خانواده محصولات زبانR ،بکارگیری R درSQL Server