

١٤٤١

لغة R



سارة باداود

محللة بيانات

١٤٤١/٩/٢٨

المحتويات

٢	مقدمة
٢	نظرة عامة حول التعلم الآلي
٣	ماهي لغة R
٣	نظرة تاريخية حول لغة R
٣	المميزات الأساسية للغة R
٣	بعض تطبيقات لغة R في حياتنا
٤	كيفية الدخول إلى عالم تحليل البيانات باستخدام لغة R
٥	بعض المراجع
٥	كتب مجانية
٥	دورات
٥	مدونات
٦	مدونة عربية
٦	مواقع
٦	مقالات
٦	تطبيق ما تعلمته
٧	المراجع

مقدمة

مع انتشار الذكاء الاصطناعي ، والتعلم الآلي ، والتعلم العميق و التحليلات المتقدمة أصبح لدى الأشخاص المهتمين بمعرفة المزيد عن هذه المجالات العديد من الأسئلة التي ينبغي مراعاتها. في هذه المقالة سيتم التطرق إلى التعلم الآلي حيث يعتبر قسم فرعي من الذكاء الاصطناعي ويعنى بالأجهزة التي يمكن للأنظمة من خلالها التعلم تلقائيا والتحسين من التجربة. وسيتم التركيز على لغة R والتي تعد واحدة من اللغات المستخدمة في التعلم الآلي.

نظرة عامة حول التعلم الآلي

لتعريف التعلم الآلي بشكل بسيط يمكن القول أنه العلم الذي يجعل الآلات تتعلم وتتصرف بطريقة مماثلة للبشر بينما نتعلم أيضا بشكل مستقل من التفاعلات الواقعية ومجموعة بيانات التدريب التي نطعمها لها.

التعلم الآلي ليس تقنية جديدة ، الخوارزميات التي تقود للتعرف على الأنماط وتطبيقات التعلم الآلي موجودة منذ سنوات عديدة. ومع ذلك ، بدأت نماذج التعلم الآلي الآن فقط في التفاعل مع مجموعات بيانات أكثر تعقيدا والتعلم من الحسابات والتنبؤات السابقة لإنتاج قرارات ونتائج موثوقة. حتى نستطيع تعليم الآلة نحتاج إلى ثلاثة مكونات:

١. مجموعة بيانات (Dataset). يمكن أن تشمل أرقام أو صور أو نصوص أو أي نوع آخر من البيانات.
٢. الميزات أو السمات (Features). الميزات هي أجزاء مهمة من البيانات تعمل كمفتاح لحل المهمة ، توضح للآلة ما يجب الانتباه إليه. ويجدر التنبيه إلى أن هذه الخاصية تختلف في التعلم الخاضع للإشراف (Supervised Learning) عن التعلم الغير خاضع للإشراف (Unsupervised Learning).
٣. الخوارزمية (Algorithm). من الممكن حل نفس المشكلة باستخدام خوارزميات مختلفة. اعتمادا على الخوارزمية ، يمكن أن تكون دقة أو سرعة الحصول على النتائج مختلفة.

بناء على كيفية تعلمهم ، تنقسم خوارزميات التعلم الآلي إلى أربع مجموعات:

١. التعلم الخاضع للإشراف (Supervised Learning). تعني كلمة "خاضع للإشراف" أن عالم البيانات يساعد البرنامج طوال عملية التدريب. أي توجد بيانات تدريب مع السمات أو الصفات الخاصة بها. على سبيل المثال ، تريد تعليم الكمبيوتر وضع الكرات الحمراء والزرقاء والخضراء في سلال مختلفة.
٢. التعلم الغير خاضع للإشراف (Unsupervised Learning). ونعني بذلك أنك لا تقدم أي ميزات أو سمات للبرنامج مما يسمح له بالبحث عن الأنماط بشكل مستقل. وغالبا ما يستخدم التعلم الغير خاضع للإشراف لتقسيم البيانات إلى مجموعات حسب التشابه. على سبيل المثال ، لديك سلة كبيرة من الغسيل يجب على الكمبيوتر فصلها إلى فئات مختلفة: الجوارب و القمصان و الجينز.
٣. التعلم شبه الخاضع للإشراف (Semi-Supervised Learning). ونعني بالشبه خاضع للإشراف أن بيانات الإدخال عبارة عن مزيج من العينات المصنفة (Labeled) والغير مصنفة (Unlabeled). ومن تطبيقاتها: تحليل الكلام (Speech Analysis) نظرا إلى أن تسمية الملفات الصوتية مهمة مكثفة للغاية ، فإن التعلم شبه الخاضع للإشراف هو نهج ممتاز لحل هذه المشكلة.
٤. التعلم التعزيزي (Reinforcement Learning). يعتبر مشابه جدا لكيفية تعلم البشر من خلال التجربة. لا يحتاج البشر إلى إشراف مستمر للتعلم بفاعلية كما في التعلم الخاضع للإشراف. على سبيل المثال ، يتعلم الطفل عدم لمس المقلاة الساخنة بعد الشعور بالألم.

ماهي لغة R

يمكن القول أن لغة R امتداد للغة S.

نظرة تاريخية حول لغة R

قبل أن نتطرق إلى لغة R من المهم فهم فلسفة لغة S العامة لمستخدمي لغة S ولغة R لأنها تمهد الطريق لتصميم اللغة نفسها ، من المهم أن ندرك أن لغة S لها جذورها في تحليل البيانات ، ولم تأتي من خلفية لغة برمجة تقليدية بل كان مخترعوها يركزون على اكتشاف كيفية جعل تحليل البيانات أسهل ، لأنفسهم أولا ، ومن ثم للآخرين.

ولنبدأ في عام ١٩٧٦ تم تطوير لغة تدعى بلغة S كبيئة للتحليل الإحصائي. في عام ١٩٨٨ تمت إعادة كتابة النظام بلغة C وبدأت تشبه النظام الذي لدينا اليوم. منذ أوائل التسعينات ، سارت حياة لغة S في مسار متعرج إلى حد ما. فقد كانت أحد القيود الرئيسية للغة S هو أنها متوفرة فقط في حزمة تجارية ، S-PLUS. في عام ١٩٩١ ، تم إنشاء لغة R من قبل Ross Ihaka and Robert Gentleman في قسم الإحصاء في جامعة أوكلاند. في عام ١٩٩٣ تم الإعلان عن لغة R لأول مرة للجمهور. تم توثيق تجربة Ross Ihaka and Robert Gentleman في تطوير لغة R في ورقة علمية عام ١٩٩٦ في مجلة Computational and Graphical Statistics.

المميزات الأساسية للغة R

١. تعمل لغة R على أي نظام تشغيل.
٢. مفتوحة المصدر ، تعني أن أي شخص حر في تكييف البرنامج مع أي منصة يختارها.
٣. الإصدارات المتكررة. تشير الإصدارات المتكررة إلى التطور النشط للبرنامج وتتضمن معالجة الأخطاء في الوقت المناسب.
٤. قدراتها الرسومية المتطورة. ومن أمثلة هذه الحزم حزمة (ggplot2).

بعض تطبيقات لغة R في حياتنا

١. البحوث الأكاديمية. R هي أداة بحث إحصائية ، لاتزال تستخدم من قبل الإحصائيين والطلاب والباحثين من أجل إجراء حسابات وتحليلات إحصائية مختلفة. وتستخدم أيضا في الأبحاث التي تتضمن مجموعات كبيرة من البيانات مثل البيانات الضخمة ، وإيجاد الشذوذ والأنماط الجينية ، والتركيبات الدوائية المختلفة. ومن الأمثلة على ذلك ، جامعة كورنيل (Cornell University): حيث يوصي كورنيل الباحثين والطلاب باستخدام لغة R لجميع أبحاثهم التي تنطوي على الحوسبة الإحصائية وجامعة كاليفورنيا أيضا (University of California).
٢. قطاع تكنولوجيا المعلومات. لا تستخدم شركات تكنولوجيا المعلومات لغة R فقط لذكاء الأعمال الخاص بها ، ولكنها تقدم مثل هذه الخدمات أيضا للشركات الصغيرة والمتوسطة والكبيرة. تستخدم أيضا لمنتجات التعلم الآلي. ومن أمثلة شركات تكنولوجيا المعلومات الكبرى التي تستخدم لغة R شركة Accenture و شركة IBM وشركة Infosys وغيرها.
٣. العلوم المالية. تقدم لغة R مجموعة مكتبات إحصائية متقدمة لجميع المهام والحسابات المالية. مثل تحليل السلاسل الزمنية ، استخراج البيانات المالية ، تقييم المخاطر السلبية وغيرها. ومن أمثلة استخدامها في هذا القطاع Lloyds of London يستخدم هذا القطاع لغة R لتحليل المخاطر.
٤. التجارة الإلكترونية. في مجال التمويل والبيع بالتجزئة ، تعد التحليلات مفيدة لتقييم المخاطر ووضع استراتيجية تسويقية. تستخدم شركات التجارة الإلكترونية لغة R لتحسين تجربة المستخدم على موقعها وكذلك لأغراض التسويق والتمويل. ومن الأمثلة على ذلك أمازون (Amazon) تستخدم أمازون لغة R لتحليل البيانات وتحسين الاقتراحات عبر المنتجات.

٥. الرعاية الصحية. يتم استخدام لغة R في التحليل والتنبؤ بانتشار الأمراض المختلفة ، وتحليل التسلسلات الجينية و بيانات سلامة الأدوية وغيرها. توفر حزم R أيضا تسهيلات لتحليل البيانات الجينومية.
٦. التصنيع. تستخدم شركات التصنيع المختلفة لغة R لتكامل استراتيجياتها التسويقية والتجارية. يحلون تعليقات العملاء للمساعدة في تبسيط وتحسين منتجاتهم. ومن الأمثلة على ذلك ، شركة فورد للسيارات (Ford Motor Company) تستخدم فورد لغة R في التحليلات الإحصائية لدعم استراتيجيتها في العمل وتحليل مشاعر العملاء حول منتجها مما يساعدهم في تحسين تصميماتهم المستقبلية.

مما سبق ذكره يمكن أن نعرف لغة R بأنها لغة برمجية مفتوحة المصدر وبيئة شائعة الاستخدام في تحليل البيانات وتصويرها وتعد واحدة من أكثر اللغات المستخدمة بين الإحصائيين.

كيفية الدخول إلى عالم تحليل البيانات باستخدام لغة R

R هي لغة برمجة شائعة بشكل متزايد ، خاصة في عالم تحليل البيانات وعلوم البيانات. لكن تعلم لغة R يمكن أن يكون تحديًا محبطًا إذا لم تكن متأكدًا من كيفية التعامل معه. لذا سأذكر هنا بعض الخطوات التي يمكن أن تساعد من يرغب في تعلم تحليل البيانات باستخدام لغة R :

١. قبل الدخول إلى عالم تحليل البيانات باستخدام هذه اللغة ينبغي عليك الإجابة على هذا السؤال : لماذا تستخدم لغة R ؟ أو كيف ستكون لغة R مفيدة لي ؟.
٢. إعداد جهازك لتحميل لغة R.
٣. تعلم أساسيات لغة R. يجب أن تبدأ بفهم أساسيات اللغة والمكتبات وهيكل البيانات. تعلم برمجة R ومعالجة البيانات.
٤. فهم مجتمع R. السبب الرئيسي للنمو السريع لـ R ونجاحه الهائل هو مجتمعه القوي.
٥. استيراد البيانات ومعالجتها. يعد استيراد البيانات ومعالجتها أحد الخطوات المهمة في سير عمل عالم البيانات ، يسمح R باستيراد تنسيقات بيانات مختلفة باستخدام حزم محددة يمكن أن تجعل عملك أسهل.
٦. التصوير الفعال للبيانات. يقدم R طرقًا متعددة لإنشاء الرسوم البيانية. الطريقة القياسية هي من خلال الاستفادة من الرسومات الأساسية في R. ومع ذلك ، هناك أدوات أفضل (أو حزم) لإنشاء الرسوم البيانية الخاصة بك بطريقة أكثر بساطة وأكثر جمالًا مثل حزم (ggplot2).
٧. تعلم كلا من التنقيب عن البيانات (Data Mining) و التعلم الآلي (Machine Learning).
٨. كتابة النتائج باستخدام لغة R. إن مشاركة نتائجك مع زملائك المتحمسين لعلوم البيانات مهم جدًا بقدر أهمية التحليل نفسه. لحسن الحظ R لديه بعض الأدوات الرائعة للقيام بذلك والتي يمكن أن توفر لك الكثير من الوقت ، مثل R Markdown.
٩. الممارسة ثم الممارسة. سوف تصبح مبرمج R عظيم فقط من خلال الممارسة. لذلك ، تأكد من معالجة تحديات علوم البيانات الجديدة بانتظام.

بعض المراجع

كتب مجانية

١. اسم الكتاب (An Introduction to R). رابط تحميل الكتاب
(<https://cran.r-project.org/doc/manuals/R-intro.pdf>)
٢. اسم الكتاب (R Programming for Data Science). رابط تحميل الكتاب
(<https://bookdown.org/rdpeng/rprogdatascience/>)
٣. اسم الكتاب (R Graphics Cookbook). رابط تحميل الكتاب
(<http://www.cookbook-r.com/>)
٤. اسم الكتاب (Data Visualization - A Practical Introduction). رابط تحميل الكتاب
(<https://socviz.co/>)
٥. اسم الكتاب (The Essentials of Data Science). رابط الكتاب
(<https://togaware.com/onepager/>)

دورات

١. اسم الدورة (R Programming). رابط الدورة
(<https://www.coursera.org/learn/r-programming>)
٢. اسم الدورة (R Programming Courses). رابط الدورة
(<https://www.edx.org/learn/r-programming>)
٣. اسم الدورة (Learn R). رابط الدورة
(<https://www.codecademy.com/learn/learn-r>)
٤. اسم الدورة (R for Data Analysis). رابط الدورة
(<https://alison.com/course/r-for-data-analysis>)
٥. اسم الدورة (Kaggle R Tutorial on Machine Learning). رابط الدورة
(<https://www.datacamp.com/community/open-courses/kaggle-r-tutorial-on-machine-learning>)
٦. اسم الدورة (R Basics – R Programming Language Introduction). رابط الدورة
(<https://www.udemy.com/course/r-basics/>)

مدونات

١. اسم المدونة (Revolutions). رابط المدونة
(<https://blog.revolutionanalytics.com/>)
٢. اسم المدونة (Data Science Central). رابط المدونة
(<https://www.datasciencecentral.com/>)
٣. اسم المدونة (R-bloggers). رابط المدونة
(<https://www.r-bloggers.com/>)
٤. اسم المدونة (Big Data Made Simple). رابط المدونة
(<https://bigdata-madesimple.com/>)
٥. اسم المدونة (Probably Overthinking It). رابط المدونة
(<https://365datascience.com/51-data-science-blogs/>)
٦. اسم المدونة (Oracle and Data Science). رابط المدونة
(<https://blogs.oracle.com/datascience/>)

٧. اسم المدونة (Airbnb Data blog). رابط المدونة
(<https://365datascience.com/51-data-science-blogs/>).
مدونة عربية

١. اسم المدونة (مدونة عزيز). رابط المدونة
(<https://www.aziz-blog.com/?p=948>).

مواقع

١. اسم الموقع (Tutorials point). رابط الموقع
(<https://www.tutorialspoint.com/r/index.htm>).
٢. اسم الموقع (Quick-R). رابط الموقع
(<https://www.statmethods.net/r-tutorial/index.html>).
٣. اسم الموقع (Data Mentor). رابط الموقع
(<https://www.datamentor.io/r-programming/>).
٤. اسم الموقع (Guru99). رابط الموقع
(<https://www.guru99.com/r-tutorial.html>).
٥. اسم الموقع (W3schools). رابط الموقع
(<https://www.w3schools.in/r/>).
٦. اسم الموقع (ggplot2). رابط الموقع
(<https://ggplot2.tidyverse.org/>).

مقالات

١. اسم المقالة (R – Statistical Programming Language). رابط المقالة
(<https://towardsdatascience.com/r-statistical-programming-language-6adc8c0a6e3d>).
٢. اسم المقالة (Why R? The Pros and Cons of the R Language). رابط المقالة
(<https://www.infoworld.com/article/2940864/r-programming-language-statistical-data-analysis.html>).
٣. اسم المقالة (R Programming Language). رابط المقالة
(<https://radiopaedia.org/articles/r-programming-language>).

تطبيق ما تعلمته

١. <https://www.kaggle.com>
٢. <https://datahack.analyticsvidhya.com/contest/all/>

- [1] Peng, R. D. (2016). *R Programming for data science*. Victoria, British Columbia, Canada: Leanpub.
- [2] Brownlee, J. (2019, October 7). Your First Machine Learning Project in R Step-By-Step. Retrieved from <https://machinelearningmastery.com/machine-learning-in-r-step-by-step/>
- [3] Burger, S. (2018). *Introduction to machine learning with R: rigorous mathematical analysis*. O'Reilly Media, Inc.
- [4] What Is Machine Learning?: How It Works, Techniques & Applications. (n.d.). Retrieved from <https://www.mathworks.com/discovery/machine-learning.html>
- [5] Heidenreich, H. (2018, December 4). What are the types of machine learning? Retrieved from <https://towardsdatascience.com/what-are-the-types-of-machine-learning-e2b9e5d1756f>
- [6] Potentia Analytics. (2019, December 18). Potentia Analytics. Retrieved from <https://www.potentiaco.com/what-is-machine-learning-definition-types-applications-and-examples/>
- [7] What is R? (n.d.). Retrieved from <https://www.r-project.org/about.html>
- [8] S-Plus Programming: How To Get Started With Statistical Programming. (2018, February 13). Retrieved from <https://www.whoishostingthis.com/resources/s-plus/>
- [9] Gour, R. (2019, April 25). A Complete Guide to Learn R. Retrieved from <https://towardsdatascience.com/a-complete-guide-to-learn-r-29e691c61d1>
- [10] Agrawal, V. (2017, August 3). Applications Of R Programming In R-eal World. Retrieved from <https://elearningindustry.com/applications-r-programming-r-eal-world>
- [11] Team, T. V. (2019, December 29). R Applications - 9 Real-world Use Cases of R programming. Retrieved from <https://techvidvan.com/tutorials/r-applications/>