

Machine Learning

Guías de Estudio Semanales

Comentario general sobre los contenidos de la primera unidad

Por una parte se introduce los conceptos básicos de Machine Learning (ML) que se tiene que tener siempre presente para cada algoritmo estudiado. Cabe destacar los conceptos siguientes:

- Los datos disponibles
- Abstracción
- Generalización
- Evaluación del modelo

que en la práctica de cualquier algoritmo de ML se puede expresar como:

- Recopilación de los datos disponibles
- Exploración y preparación de los datos
- Entrenamiento del modelo
- Evaluación del modelo
- Mejorar el modelo

Por otra parte se presenta el concepto de *informe dinámico* o *investigación reproducible* que se usará durante todo el curso como una forma adecuada de presentar los informes y, en un sentido más amplio, este concepto puede salir del marco académico y llegar al trabajo diario como una forma de ser más productivos.

Los contenidos de la unidad son:

- 1.1. Los orígenes del Machine Learning.
- 1.2. ¿Cómo aprenden las máquinas?
- 1.3. Pasos en la implementación de Machine Learning.
- 1.4. La elección del algoritmo de Machine Learning
- 1.5. Elementos básicos de R.
- 1.6. Machine Learning con R e informes dinámicos.

El primer bloque (1.1-1.5) hace un repaso histórico del ML, plantea sus límites e indica los pasos necesarios que realiza los métodos de ML.

El segundo bloque (1.6-1.7) tiene dos partes, la primera es un repaso de los objetos de R y de la manipulación de los datos que ya habéis visto y trabajado en asignaturas precedentes. En cambio, la segunda puede ser novedosa, aparece el concepto de **informe dinámico** o **investigación reproducible**. El objetivo es tener un fichero ejecutable que incluya texto y código que al compilarse de como resultado final un informe en html, pdf, ... que incluya el texto y los resultados del análisis. Si este informe dinámico se hace lo suficientemente general, con cambiar el fichero de datos se puede tener un nuevo informe que se adapte a los nuevos datos. Por otra parte, es una manera muy práctica de mostrar a otros investigadores el trabajo que has realizado ya que el investigador puede reproducirlo.

En nuestro caso, se harán informes reproducibles del tipo R markdown. Esto quiere decir que el texto se hace en markdown (lenguaje muy sencillo que usa tags) y R para como lenguaje de programación. Otra opción posible, más complicada, es usar Latex como lenguaje de texto en lugar de markdown. En este caso, se tiene toda la potencia de Latex en la presentación de resultados.

Materiales de estudio

El material del básico pero no necesariamente exclusivo es el libro de referencia Lantz – 15. Este libro mezcla de una manera muy asequible los conceptos y la implementación de un ejemplo guía en R en cada algoritmo que explica.

En cuanto al material para estudiar los informes dinámicos en markdown se tiene:

- <http://rmarkdown.rstudio.com/lesson-1.html>
- <https://bookdown.org/yihui/rmarkdown/>
- <https://bookdown.org/yihui/bookdown/>
- <http://yihui.name/slides/2012-knitr-RStudio.html>
- <https://www.rstudio.com/resources/cheatsheets/>
- <https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf>

El primer link seria el básico donde se puede aprender y ver todo el potencial que tiene. Es recomendable, pero no es obligatorio, hacer servir Rstudio como editor de R y markdown.

Un primer ejemplo de informe dinámico con R markdown desde Rstudio se puede conseguir al crear un fichero R markdown. Los pasos son File/New File/R Markdown .../ que presenta un cuadro de dialogo que se debe completar o dejar por defecto y después del “ok” se tiene un documento R markdown. Para ejecutar el script y obtener un informe en formato html solo hay que hacer “click” al botón knit HTML.

A título informativo, en este link del repositorio de R, <https://cran.r-project.org/web/views/ReproducibleResearch.html>, se puede ver la cantidad de packages que se tiene disponible en la materia de informe dinámicos en R.

Unidad \ Fuente	Lantz – 15
1.1. Los orígenes del Machine Learning.	Cap. 1 (hasta "Machine learning ethics")
1.2. ¿Cómo aprenden las máquinas?	Cap. 1 ("How machines learn")
1.3. Pasos en la implementación de Machine Learning.	Cap. 1 ("Machine learning in practice" and "Types of input data")
1.4. La elección del algoritmo de Machine Learning	Cap. 1 (de "Types of machine learning algorithms" hasta "Summary")
1.5. Elementos básicos de R.	Cap 2
1.6. Machine Learning con R e informes dinámicos.	http://rmarkdown.rstudio.com/lesson-1.html https://bookdown.org/yihui/rmarkdown/ https://bookdown.org/yihui/bookdown/ http://yihui.name/slides/2012-knitr-RStudio.html https://www.rstudio.com/resources/cheatsheets/ https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf

Actividades complementarias

Las actividades prácticas que podéis realizar para reforzar el aprendizaje son:

- Lectura de otras fuentes de información.
- Realización de algunos informes dinámicos con códigos de R previos.
- Trabajar los conceptos con ordenador (por ejemplo usando R).

Si encuentras materiales complementarios de interés o que desees discutir puedes añadirlos a la actividad **"Upload Material añadido por el estudiante para compartir"** en Moodle 2. La idea es muy simple, al entrar en la actividad aparecen diferentes pestañas. Si se tiene un pdf, video, url, ... que quieres subir hay que ir a la pestaña añadir entrada y completar la ficha. Una vez completa hacer clic al botón inferior. Si se quieren ver las aportaciones actuales hay que ir Ver lista.