



UNIVERSIDAD
COMPLUTENSE
MADRID

Proyecto de innovación
Convocatoria 2022/2023

Nº de proyecto: 202

Docencia de medidas de divergencia en el entorno virtual Moodle a partir de ejercicios
resueltos

Responsable del proyecto: Pedro Miranda Menéndez

Facultad de Ciencias Matemáticas
Departamento: Estadística e Investigación Operativa

1.- Objetivos propuestos en la presentación del proyecto.

El objetivo de este proyecto se engloba en la línea

Innovación en recursos educativos en abierto y enseñanza virtual.

El objetivo principal de este proyecto era programar una aplicación que permitiese a los alumnos practicar las herramientas teóricas vistas en clase de Doctorado. Estaba planteado como un primer paso para ir generando programas que generen problemas de forma aleatoria en los que aplicar distintos tipos de aplicaciones de las medidas de divergencia estudiadas en las clases teóricas del curso de doctorado. En este caso nos planteamos generar programas aplicados a la resolución del Análisis de Correlaciones Canónicas.

El Análisis de Correlaciones Canónicas clásico sólo es capaz de encontrar relaciones entre los vectores aleatorios de tipo lineal; sin embargo, en muchas ocasiones existen relaciones de otros tipos que son interesantes de estudiar. Recientemente se han propuesto alternativas al Análisis de Correlaciones Canónicas basados en medidas de divergencia que son capaces de detectar estas relaciones no lineales.

La idea teórica es la misma que en el análisis clásico pero cambian las medidas de distancia entre los vectores aleatorios. Se trata también de que los alumnos se puedan familiarizar con estas nuevas técnicas a nivel de aplicación práctica.

En definitiva, se trataba de generar datos aleatorios a partir de distintos modelos, de forma que los alumnos tubiesen que responder a varias preguntas sobre las salidas de ordenador al analizar estos datos, tanto para el caso clásico como para las técnicas más modernas basadas en otras medidas de divergencia.

Para ello, se ha usado el entorno exams del lenguaje de programación R, que permiten la inclusión de los programas en el campus virtual, además de ser un software libre. La bibliografía fundamental en la que nos hemos basado es

- **Flexible Generation of E-Learning Exams in R: Moodle Quizzes, OLAT Assessments, and Beyond.** Zeileis, N. Umlauf, F. Leisch. *Journal of Statistical Software*, 58 (1), (2014).

Otra posibilidad de uso de esta aplicación consiste en poder evaluar de forma online los contenidos de la asignatura. Pero este objetivo no se planteó en la propuesta original, ya que el objetivo primordial es que los alumnos puedan autoevaluarse y ver si han asimilado correctamente los algoritmos.

2.- Objetivos alcanzados

Se han desarrollado varios programas (en concreto 4) que generan los datos de problemas de Análisis de Correspondencias Canónicas en distintas situaciones de forma aleatoria.

Los ficheros que se han desarrollado son:

- Análisis de Correspondencias Canónicas cuando las componentes son de tipo lineal sin contaminación en los datos.
- Análisis de Correspondencias Canónicas cuando las componentes son de tipo lineal con contaminación en los datos.
- Análisis de Correspondencias Canónicas cuando las componentes no son de tipo lineal sin contaminación en los datos.
- Análisis de Correspondencias Canónicas cuando las componentes no son de tipo lineal con contaminación en los datos.

Estos cuatro problemas reflejan las cuatro situaciones posibles en las que se ve la utilidad de las medidas de divergencia:

- En el primer caso, es de esperar un buen funcionamiento de las herramientas clásicas, del modelo de Yin y de las medidas de divergencia.
- En el tercer caso, se espera un buen funcionamiento del modelo desarrollado por Yin y de las medidas de divergencia, pero no de las técnicas clásicas.
- En los casos segundo y cuarto, es de esperar un buen funcionamiento de las medidas de divergencia robustas, pero no de las herramientas clásicas ni de las desarrolladas por Yin.

3.- Metodología empleada en el proyecto

El proyecto se estructuraba en dos fases diferenciadas.

1. **FASE 1:** Construir la aplicación que permitiese generar los problemas de los distintos algoritmos de forma aleatoria. Esta parte se había planteado para ser desarrollada antes del comienzo de las clases del curso de doctorado. En esta fase se han fijado también la forma de plantear el problema. Se ha decidido para no desanimar a los alumnos y ya que el objetivo es practicar el funcionamiento de los algoritmos que el enunciado sea lo más sencillo posible y que el número de preguntas sea lo más reducido posible.
2. **FASE 2:** Probar y corregir posibles problemas que pudiesen aparecer en el uso práctico de la aplicación, tanto desde el punto de vista computacional como de interpretación de los enunciados. Esta parte estaba planteada para ser realizada en el momento en el que los alumnos pudiesen probar los programas y sugerir mejoras o detectar errores que se pudieran haber pasado por alto en la primera etapa. Sin embargo, el curso de doctorado no ha sido impartido este curso, con lo que esta segunda fase no ha podido ser aplicada.

La metodología que se ha seguido con este proyecto era permitir a los alumnos practicar estos algoritmos de forma autónoma y dedicando el tiempo que cada uno de ellos considerase necesario. En ningún momento se planteó acceder a las respuestas o número de ensayos que cada alumno pudiese realizar porque el objetivo era conseguir que los alumnos estuviesen cómodos y que no se sintiesen evaluados por utilizar la aplicación.

4.- Recursos humanos

El equipo que ha realizado este proyecto está formado por:

Responsable: Pedro Miranda Menéndez pmiranda@ucm.es

Equipo: Angel Felipe Ortega afelipe@ucm.es

Equipo: María Jaenada Malagón mjaenada@ucm.es

Equipo: Susana Martínez Suárez smartinezsuarez@gamil.com

Equipo: Leandro Pardo Llorente lpardo@ucm.es

5.- Desarrollo de las actividades

El proyecto se ha desarrollado según se había explicado en la solicitud. Al contrario que en el proyecto anterior, en este proyecto nos hemos encontrado con muchas menos dificultades debido a la idiosincrasia del problema a resolver.

- En primer lugar, en este caso sí hemos podido utilizar librerías en R para poder resolver los problemas. Básicamente, la dificultad en el uso de las medidas de divergencia es encontrar mínimos de funciones muy complejas. Sin embargo, los algoritmos para la resolución de estas ecuaciones ya están implementadas de manera general en R y basta llamar a las subrutinas correspondientes. De esta forma, la generación de los distintos ficheros ha sido sencilla.
- Una vez generado los distintos programas, era necesario incluirlos como ficheros de formato RnW para poder utilizarlos con la librería exams de R. Se trataba aquí de combinar los ficheros desarrollados en la etapa anterior con un fichero de tipo .tex en el que se plantean las preguntas que se hacen en cada problema y que proporciona la solución del mismo. Aunque proporcionar la solución es algo opcional en la librería exams, precisamente éste era el objetivo del proyecto, ya que no se trataba de que los alumnos practicasen, sino que pudiesen ver la solución de cada problema completamente desarrollada para comprobar que se ha asimilado el funcionamiento de cada uno de los algoritmos. Por ello, la parte de Latex del programa en RnW es especialmente larga. Un ejemplo de la vista que aparece a los alumnos en el campus virtual al utilizar la aplicación se puede ver en la Figura 2 del anexo.
- Como sucedió en el proyecto anterior, la forma visual de los problemas en formato pdf y en formato Moodle difieren y es mucho menos intuitiva en formato Moodle. Sin embargo, en este caso no es tan importante por ña forma que tiene la respuesta. Una dificultad que no habíamos previsto aparece con el entorno que proporciona Moodle para mostrar las soluciones.
- En este caso no hemos tenido problemas en el sentido de desconocer la extensión de la respuesta de cada problema. Por ello, hemos aprovechado para incluir gráficas en los ficheros generados, de forma que la idea visual de los algoritmos sea mucho más intuitiva.
- Finalmente, el entorno exams está desarrollado para ser aplicado en evaluaciones online en la que cada alumno tiene un ejercicio con datos diferentes pero de una dificultad similar. Por ello, no es posible incorporar los programas como aplicaciones en el campus virtual de la asignatura, sino que es necesario incluirlos como cuestionarios.

Hemos planteado 4 ficheros RnW diferentes, de forma que se diferencian entre sí en función del tipo de datos y de la solución del mismo, pero de forma que el alumno no sabe a priori en cuál de las cuatro situaciones se encuentra.

Otro aspecto del entorno exams que nos ha sorprendido es que esperábamos generar problemas y que la aplicación permitiese que el alumno realizase un número indefinido de problemas, de forma que en cada caso ejercicio se generasen datos de forma aleatoria. Sin embargo, el entorno exams sólo permite que el alumno realice un ejercicio seleccionado al azar de entre una batería de problemas previamente generados por la aplicación y cuyo número está prefijado en la ejecución del programa. Para solventar este problema y que los alumnos tuviesen a su disposición un número suficientemente grande de problemas, hemos generado una cantidad elevada de problemas, lo que hace que se cumplan las condiciones que se pretendían conseguir inicialmente. La instrucción de ejecución sería de la forma

```
exams2moodle("../Desktop/R/Final/Caso1.Rnw", 50)
```

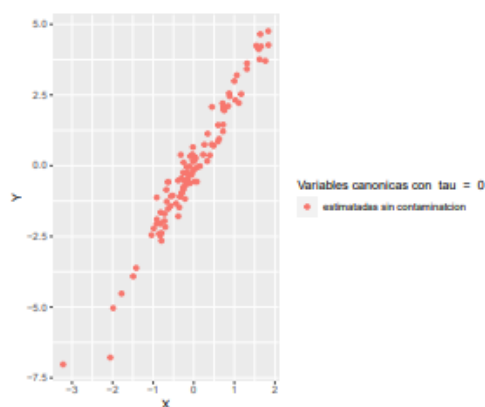
donde el 50 es el número de problemas que queremos que se generen.

La idea original del proyecto era poder utilizar estos programas en un curso de doctorado, pero este curso no se ha impartido. Esperamos poder aplicarlo en próximos cursos, en los que sí se imparta el curso de doctorado sobre medidas de divergencia.

6.- Anexos

1. Problem

Se consideran dos conjuntos de datos X e Y de dimensiones 8 y 3. Se pretende identificar las relaciones que pueda haber entre estos dos grupos de variables. Para ello, se generan los siguientes gráficos de variables canónicas estimadas con estimadores DPD con distintos valores de tau.

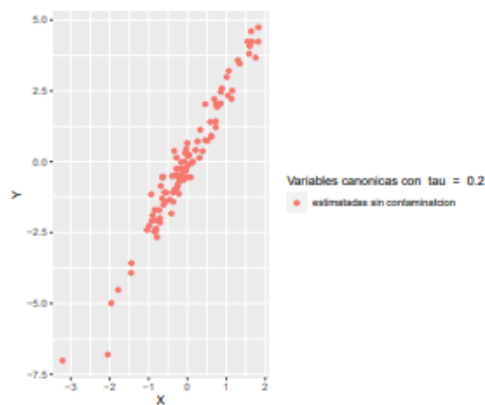


Los vectores canonicos para tau = 0 son para los dos grupos

```
[1] 0.782 0.363 0.344 -0.018 0.021 0.021 -0.038 -0.013 0.368 0.014 0.001
```

```
[1] 0.368 0.014 0.001
```

respectivamente.



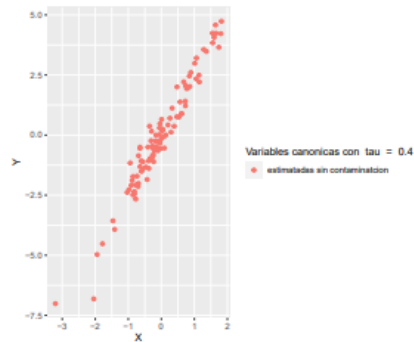
Los vectores canonicos para tau = 0.2 son para los dos grupos

```
[1] 0.786 0.359 0.338 -0.013 0.026 0.017 -0.049 -0.011
```

```
[1] 0.368 0.020 0.008
```

respectivamente.

Figura 1: Ejemplo de enunciado en la situación 1: Un problema en el que las componentes canónicas siguen un modelo lineal (Primera parte).

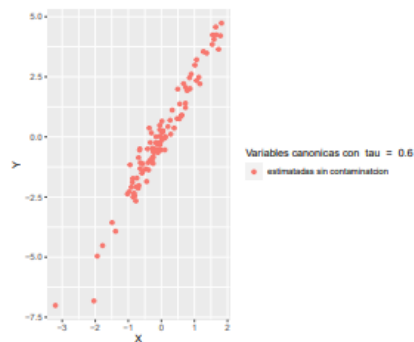


Los vectores canonicos para $\tau = 0.4$ son para los dos grupos

[1] 0.788 0.357 0.333 -0.009 0.027 0.011 -0.058 -0.009

[1] 0.368 0.023 0.011

respectivamente.



Los vectores canonicos para $\tau = 0.6$ son para los dos grupos

[1] 0.789 0.357 0.330 -0.006 0.025 0.005 -0.064 -0.007

[1] 0.368 0.025 0.012

respectivamente.

- ¿Qué tipo de relación crees que tienen?
- ¿Podría el análisis de correlaciones canónicas clásico identificar la relación?
- ¿Cuáles son los valores de los vectores canónicos en cada caso? ¿Qué variables tienen mayor relevancia en la relación?
- ¿Encuentras alguna diferencia en los distintos valores del hiperparámetro τ ?

Figura 2: Ejemplo de enunciado en la situación 1: Un problema en el que las componentes canónicas siguen un modelo lineal (segunda parte).

Solution

(a) La relación entre las variables es una relación lineal.

2

(b) Por ser la relación lineal, sí podría identificarse mediante el análisis clásico. De ser otra relación funcional cualquiera, esta técnica no valdría.

(c) Las componentes son, para $\tau = 0$: El primer grupo de variables X se combina mediante el vector a_c

```
[1] 0.789 0.357 0.330 -0.006 0.025 0.005 -0.064 -0.007
```

El segundo grupo de variables Y se combina mediante el vector b_c

```
[1] 0.368 0.014 0.001
```

El resto de los vectores se extraen de manera similar

(d) Por último, todos los estimadores encuentran relaciones similares entre los parámetros.
string Correlaciones 0.01

Figura 3: Ejemplo de solución generada por el programa para el ejemplo anterior.