

Modelos y Simulación II: Proyecto del Curso

Prof. Julián David Arias Londoño
Departamento de Ingeniería de Sistemas
Universidad de Antioquia, Colombia
julian.ariasl@udea.edu.co

Abstract—Las actividades propuestas para el desarrollo del proyecto buscan que cada uno de los grupos de estudiantes presenten todo el diseño, análisis y simulación de un sistema de predicción basado en técnicas de aprendizaje de máquina; describiendo el problema y su contexto en términos del estado del arte, especificando cada una de las etapas del desarrollo del trabajo, los modelos con sus respectivas restricciones, la metodología de validación, los resultados de las simulaciones y las conclusiones obtenidas.

Index Terms—Modelos fenomenológicos, aprendizaje de máquina.

I. EJERCICIOS

- 1) Describa de manera clara el problema de predicción que está abordando, su campo de aplicación y explique si corresponde a un problema de clasificación o de regresión.
- 2) Enumere las variables incluidas como entrada al sistema y la o las variables a predecir. Explique claramente el tipo de codificación de cada variable y, si la base de datos cuenta con valores faltantes, explique cómo se llenaron los vacíos en cada caso.
- 3) Realice una búsqueda de al menos 4 artículos que hayan abordado el mismo problema de aprendizaje que Uds están trabajando. Incluya, en la medida de lo posible, trabajos que hayan empleado la misma base de datos. Describa brevemente:
 - ¿Qué técnica(s) de aprendizaje usan en los artículos?
 - ¿Qué metodología de validación usaron?
 - ¿Cuáles fueron los resultados obtenidos en cada uno de los trabajos citados?

Se recomienda buscar trabajos en las bases de datos: www.sciencedirect.com y www.ieeexplore.org. También se pueden buscar trabajos en la base de datos <http://link.springer.com>, pero se debe tener en cuenta que el acceso que tiene la Universidad es mucho más limitado para dicha base de datos. Incluir preferiblemente artículos publicados en revista no en congresos o conferencias. No utilice más de una página del informe para esta descripción.

- 4) Incluya una sesión dentro de su informe con el nombre *Experimentos*, en la cual describa la metodología de validación usada y la base de datos que está usando para llevar a cabo el proyecto, incluyendo la fuente de la base de datos como referencia, el número de muestras, variables, etc. Si su problema es de clasificación, indique cuántas clases contiene la base de datos y cuántas muestras por clase. Si la base de datos está desbalanceada,

deben considerar el uso de técnicas de submuestreo y sobremuestreo inteligente, además de usar validación estratificada.

- 5) Si su problema es de clasificación debe evaluar el uso de los siguientes modelos de predicción y documentar los resultados de las simulaciones realizadas:
 - Naïve Bayes
 - K vecinos más cercanos
 - Redes Neuronales Artificiales
 - Random Forest
 - Máquinas de Soporte Vectorial con kernel lineal y con kernel RBF.
- 6) Si su problema es de regresión debe evaluar el uso de los siguientes modelos de predicción y documentar los resultados de las simulaciones realizadas:
 - Regresión múltiple
 - Ventana de Parzen
 - Redes Neuronales Artificiales
 - Random Forest
 - Regresión por Vectores de Soporte con kernel lineal y con kernel RBF.

Para cada modelo se deben mostrar tablas que indiquen el resultado obtenido para los diferentes parámetros evaluados. Recuerde que como parte de una correcta metodología de validación se deben incluir, según sea el caso, medidas de sensibilidad, especificidad, eficiencia, matriz de confusión, error cuadrático medio, MAPE, coeficiente de determinación, etc., Deben incluir intervalos de confianza para todas las medidas usadas.

- 7) Realice un análisis individual de cada una de las características, a partir de medidas de correlación y del índice de Fisher (según sea el caso). Identifique de acuerdo con este análisis, ¿cuáles son las características candidatas a ser eliminadas?
- 8) Realice selección de características por el método de búsqueda secuencial ascendente o descendente y evalúe nuevamente en los 3 mejores modelos evaluados. Para realizar este punto cada grupo debe decidir la función criterio a usar, justificando su decisión. Incluya una tabla con los resultados, indicando el porcentaje de reducción alcanzado y las demás medidas de desempeño. Recuerde incluir en el informe cuál fue el criterio de selección usado y porqué.
- 9) Realice extracción de características por el método PCA y evalúe nuevamente en los 3 mejores modelos de predicción evaluados. Cada grupo debe decidir el criterio para seleccionar el número de componentes justificando

su decisión. Incluya una tabla con los resultados, indicando el porcentaje de reducción alcanzado y las demás medidas de desempeño.

- 10) Incluya una sección final de discusión en la cuál analice los resultados obtenidos y los compare con los resultados de los artículos consultados en el punto 2.

Nota. El informe se debe presentar en formato IEEE (Similar al usado en esta guía). Procure realizar una redacción coherente y haga un buen uso de la referenciación (consulte las normas IEEE para incluir las citas en la bibliografía). Cada punto del informe tiene un valor diferente así:

Punto	Valor
1,2,3 y 4	1
5 ó 6	1
7	0.5
8	0.7
9	0.8
10	0.5
Sustentación	0.5
Total	5.0

Se darán puntos adicionales si el problema abordado corresponde a un paradigma de múltiples salidas o de múltiples instancias. El grupo que no presente sustentación no tendrá nota. Para la sustentación se debe preparar una presentación de 10 minutos en la que se cuente el problema abordado, las características de la base de datos y los resultados obtenidos. La nota de la sustentación es individual para cada uno de los miembros del equipo.