

Hoja de Trabajo 7.

Redes Neuronales Artificiales (RNA) y Máquinas Vectoriales de Soporte (SVM)

INTRODUCCIÓN:

Kaggle

Kaggle es una comunidad en línea de científicos de datos, propiedad de Google LLC. Permite a los usuarios encontrar y publicar conjuntos de datos, explorar y construir modelos en un entorno de ciencia de datos basado en la web, trabajar con otros científicos de datos e ingenieros de aprendizaje automático, y participar en competencias para resolver los desafíos de la ciencia de datos. Tuvo su inicio al ofrecer competencias de aprendizaje automático y ahora también ofrece una plataforma pública de datos, una mesa de trabajo basada en la nube para la ciencia de la información y educación en IA de formato corto. El 8 de marzo de 2017, Google anunció que estaban adquiriendo Kaggle.

Conjunto de datos a utilizar

<https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data>

Notas:

- La hoja de trabajo se realizará en las mismas parejas de la hoja anterior.
- Los grupos serán seleccionados por afinidad.
- La hoja no se calificará si no pertenece a ningún grupo de los creados en canvas para esta hoja.

INSTRUCCIONES

- Utilice el data set [House Prices: Advanced Regression Techniques](#) que comparte Kaggle. Debe hacer un análisis exploratorio para entender mejor los datos, sabiendo que el objetivo final es predecir los precios de las casas. Recuerde explicar bien cada uno de los hallazgos que haga. La forma más organizada de hacer un análisis exploratorio es generando ciertas preguntas de las líneas que le parece interesante investigar. Genere un informe en pdf con las explicaciones de los pasos que llevó a cabo y los resultados obtenidos. Recuerde que la investigación debe ser reproducible por lo que debe guardar el código que ha utilizado para resolver los ejercicios y/o cada uno de los pasos llevados a cabo si utiliza una herramienta visual.

ACTIVIDADES

1. Use los mismos conjuntos de entrenamiento y prueba que utilizó en las hojas anteriores.
2. Seleccione como variable respuesta la que creó con las categorías del precio de la casa.
3. Explore los datos y explique las transformaciones que debe hacerle para generar un modelo de redes neuronales.

4. Genere dos modelos de redes neuronales con diferentes topologías y funciones de activación.
5. Use los modelos para predecir el valor de la variable respuesta
6. Haga las matrices de confusión respectivas.
7. Compare los resultados obtenidos con los diferentes modelos de redes neuronales en cuanto a efectividad, tiempo de procesamiento y equivocaciones (donde el algoritmo se equivocó más, donde se equivocó menos y la importancia que tienen los errores).
8. Compare la eficiencia del mejor modelo de RNA con los resultados obtenidos con los algoritmos de las hojas de trabajo anteriores. ¿Cuál es mejor para predecir? ¿Cuál se demoró más en procesar?
9. Genere varios (más de 2) modelos de SVM con diferentes kernels y distintos valores en los parámetros c , γ y d (en caso de que utilice el polinomial).
10. Use los modelos para predecir el valor de la variable respuesta.
11. Haga las matrices de confusión respectivas.
12. Compare los resultados obtenidos con los diferentes modelos que hizo en cuanto a efectividad, tiempo de procesamiento y equivocaciones (donde el algoritmo se equivocó más, donde se equivocó menos y la importancia que tienen los errores).
13. Compare la eficiencia del mejor modelo de SVM con los resultados obtenidos con los algoritmos de las hojas de trabajo anteriores. ¿Cuál es mejor para predecir? ¿Cuál se demoró más en procesar?
14. Genere un informe de los resultados y las explicaciones.

EVALUACIÓN

- **(5 puntos)** Transformaciones y explicación de las transformaciones que se le hicieron a las variables para incluirlas en los modelos de redes neuronales.
- **(25 puntos)**. Generación de dos modelos diferentes de redes neuronales. Explicación de la topología, funciones de activación y algoritmo de entrenamiento.
- **(25 puntos)**. Generación de varios modelos diferentes de SVM de los cambios en los parámetros.
- **(6 puntos)**. Entrenamiento y predicción con los modelos generados.
- **(7 puntos)**. Generación de las matrices de confusión y explicación de los resultados obtenidos
- **(7 puntos)**. Comparación entre si los modelos de redes neuronales
- **(15 puntos)**. Comparación entre si los modelos de SVM
- **(10 puntos)**. Comparación con la efectividad de los algoritmos de las hojas de trabajo anteriores.

MATERIAL A ENTREGAR

- Archivo .r o .py con el código y hallazgos comentados
- Archivo .pdf con las conclusiones y hallazgos encontrados, (obligatorio).