

Proyecto 2. Entrega 9.

Redes Neuronales Artificiales (RNA)

INTRODUCCIÓN:

InmoValor S.A. es una empresa innovadora del sector inmobiliario que ha apostado por la transformación digital para ofrecer valoraciones precisas y objetivas de propiedades. Ante un mercado dinámico y competitivo, la compañía ha adoptado técnicas avanzadas de análisis y modelos de regresión para estimar el valor de inmuebles basándose en un amplio conjunto de datos que recopila información detallada de viviendas. Este dataset incluye variables clave como ubicación, tamaño, calidad constructiva y otros factores determinantes, lo que permite desarrollar modelos predictivos capaces de reflejar con mayor exactitud las condiciones del mercado.

La empresa ha decidido incorporar un equipo de analistas de datos con la finalidad de trabajar con el conjunto de datos "House Prices: Advanced Regression Techniques" para desarrollar modelos predictivos que proyecten de manera precisa el precio de las viviendas. Mediante el análisis de variables clave como la ubicación, el tamaño y la calidad de las propiedades, el equipo utilizará técnicas avanzadas de regresión para mejorar la estimación de valores inmobiliarios y facilitar la toma de decisiones estratégicas en el mercado de bienes raíces.

Vínculo del conjunto de datos a utilizar

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

Resultados esperados en la Séptima Entrega:

Se espera la entrega de un informe detallado donde incluya:

- Los modelos de Redes Neuronales Artificiales (RNA) para clasificación usando la variable categórica de los precios que creó.
- Los modelos de RNA para predicción de los precios de las casas.
- Una comparación para estimar los precios de las viviendas donde determine qué algoritmo funcionó mejor (Árboles de regresión, Random Forest, Bayes Ingenuo, KNN, Regresión Lineal, SVR, RNA). Compare los resultados del mejor modelo de cada uno de los algoritmos. Recuerde que, para establecer una comparación válida, es necesario comparar bajo las mismas condiciones. Mismo conjunto de datos de entrenamiento y de prueba.
- Una comparación para estimar la categoría de los precios de las viviendas donde determine qué algoritmo funcionó mejor (Árboles de decisión, Random Forest, Bayes Ingenuo, KNN, Regresión Logística, SVM, RNA). Compare los resultados del mejor modelo de cada uno de los algoritmos. Recuerde que para establecer una comparación válida, es necesario comparar bajo las mismas condiciones. Mismo conjunto de datos de entrenamiento y de prueba.
- Al ser esta la última entrega de la consultoría InmoValor le ha solicitado que llegue a conclusiones sobre cuál es el algoritmo que funciona mejor con este conjunto de datos.

Debe hacer esto tanto para modelos de clasificación, con la variable categórica de precios como para los modelos de regresión que permitan predecir el precio de las casas.

- Como la vez anterior, le han pedido un informe formal, que enlace los temas usando subtítulos explicativos que mantenga al lector en un hilo conductual de los temas que aborda la consultoría. No incluya los enunciados de las actividades de esta guía como subtítulos.

ACTIVIDADES

1. Use los mismos conjuntos de entrenamiento y prueba que utilizó en las entregas anteriores.
2. Seleccione como variable respuesta la que creó con las categorías del precio de la casa.
3. Genere dos modelos de redes neuronales que sean capaz de clasificar usando la variable respuesta que categoriza las casas en baratas, medias y caras. Estos modelos deben tener diferentes topologías y funciones de activación.
4. Use los modelos para predecir el valor de la variable respuesta.
5. Haga las matrices de confusión respectivas.
6. Compare los resultados obtenidos con los diferentes modelos de clasificación **usando 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).
7. Analice si no hay sobreajuste en los modelos.
8. Para el modelo elegido de clasificación tune los parámetros y discuta si puede mejorar todavía el modelo sin llegar a sobre ajustarlo.
9. Seleccione ahora el SalesPrice como variable respuesta.
10. Genere dos modelos de regresión con redes neuronales con diferentes topologías y funciones de activación para predecir el precio de las casas.
11. Compare los dos modelos de regresión y determine cuál funcionó mejor para predecir el precio de las casas.
12. Analice si no hay sobreajuste en los modelos. Use para esto la curva de aprendizaje.
13. Para el modelo elegido de regresión tune los parámetros y discuta si puede mejorar todavía el modelo sin llegar a sobre ajustarlo.
14. Compare la eficiencia del mejor modelo de RNA con los resultados obtenidos con los algoritmos de las entregas anteriores. ¿Cuál es mejor para predecir? ¿Cuál se demoró más en procesar?
15. Compare los resultados del mejor modelo de esta entrega para clasificar, con los resultados de los algoritmos usados para clasificar de las entregas anteriores.
16. Compare los resultados del mejor modelo para predecir el precio de venta con los resultados de los algoritmos usados para el mismo propósito de las entregas anteriores.
17. Ahora que ha usado todos los modelos que hemos visto y aplicados al conjunto de datos llegue a conclusiones sobre cual es o cuales son los mejores modelos para clasificar dadas las características del conjunto de datos. ¿Cuál o cuáles son los mejores para predecir el precio de las casas? Elabore una tabla de resumen con las métricas de los modelos que está comparando.

18. Genere un informe total de la consultoría donde incluya todas las entregas parciales. Debe incluir desde el análisis exploratorio hasta esta entrega. Debe ser un informe general completamente coherente, sin subtítulos relacionados con las instrucciones de las actividades de las entregas. Debe ser un informe formal, que pueda ser presentado a los directivos de la compañía.

EVALUACIÓN

Notas:

- Tiene que poderse comprobar su aporte al trabajo grupal a través de commits. Si no existen al menos 3 commits con su aporte significativo no va a tener nota de la hoja de trabajo. Utilice una herramienta que permita registrar los aportes de cada uno.
- Debe entregar los avances durante el período de clase para poder tener derecho a la calificación de la entrega.
- **(10 puntos)** Creación de los dos modelos de clasificación usando diferentes topologías y funciones de activación. Queda claro cuales fueron las topologías y funciones de activación usadas en cada modelo.
- **(10 puntos)**. Matriz de confusión de ambos modelos para clasificar.
- **(15 puntos)**. Comparación de resultados y selección del mejor modelo de clasificación. Da una explicación amplia de sus resultados. Analiza si no tienen sobreajuste y tunea los parámetros del mejor modelo.
- **(10 puntos)**. Creación de los dos modelos de regresión usando diferentes topologías y funciones de activación. Queda claro cuáles fueron las topologías y funciones de activación usadas en cada modelo.
- **(15 puntos)**. Comparación de resultados y selección del mejor modelo de regresión. Da una explicación amplia de sus resultados. Analiza si no tienen sobreajuste y tunea los parámetros del mejor modelo.
- **(12 puntos)**. Comparación con la efectividad de los algoritmos de las hojas de trabajo anteriores.
- **(18 puntos)**. Informe general de la consultoría. Elabora un reporte formal que incluye todas las fases del proyecto, desde análisis exploratorio hasta esta entrega. La división de secciones es coherente con el tema que trata, no hay enunciados de las instrucciones de cada entrega. Está listo para presentárselo a los directivos de la empresa.
- **(10 puntos)**. Conclusiones. Debe seleccionar los mejores modelos de clasificación y regresión para el conjunto de datos.

MATERIAL A ENTREGAR

- Archivo .r o .py con el código y hallazgos comentados
- Link de Google docs con las conclusiones y hallazgos encontrados. Puede usar también Jupyter Notebooks o rmd.
- Vínculo del repositorio usado para trabajar la hoja de trabajo.

FECHAS DE ENTREGA

- **AVANCE:** Puntos del 1 al 8 de la sección de actividades: viernes 2 de mayo a las 23:59.
- **ENTREGA FINAL:** lunes, 4 de mayo a las 23:59