

## 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.

##### Competencias de Kaggle

Desde sus inicios se ha dedicado a organizar competencias para estimular a los científicos de datos a resolver problemas reales sirviendo a grandes empresas y a organizaciones sociales.

##### Opciones de competencias para esta hoja de trabajo

1. PetFinder.my Adoption Prediction. How cute is that doggy in the shelter? (<https://www.kaggle.com/c/petfinder-adoption-prediction>) *“En esta competencia, desarrollarás algoritmos para predecir la capacidad de adopción de las mascotas, específicamente, ¿con qué rapidez se adopta una mascota? Si tienen éxito, se adaptarán a las herramientas de inteligencia artificial que guiarán a los refugios y rescatadores de todo el mundo para mejorar el atractivo de los perfiles de sus mascotas, reducir el sufrimiento y la eutanasia de los animales.”*

##### Notas:

- La hoja de trabajo se realizará en parejas.
- 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.

#### ACTIVIDADES

1. Use los mismos conjuntos de entrenamiento y prueba para probar el algoritmo.
2. Explore los datos y explique las transformaciones que debe hacerle para generar un modelo de redes neuronales.
3. Especifique cuál va a ser la variable respuesta, ¿qué transformaciones tuvo que hacerle?
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 el resultado obtenido 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  $c$ ,  $\gamma$  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 el resultado obtenido 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. Actualice el kernel en kaggle donde ponga el código generado en esta tarea para que compita por uno de los premios.
15. Genere un informe de los resultados y las explicaciones, tanto en html, o pdf.

## EVALUACIÓN

- **(5 puntos)** Transformaciones y explicación de las transformaciones que se le hicieron a las variables para incluirlas en los modelos de redes neuronales.
- **(10 puntos)**. Selección de la variable respuesta y explicación de las transformaciones realizadas.
- **(20 puntos)**. Generación de dos modelos diferentes de redes neuronales. Explicación de la topología, funciones de activación y algoritmo de entrenamiento.
- **(20 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).
- Link del kernel creado en kaggle.