

## Hoja de Trabajo 8. 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 parejas o tríos.
- 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 máquinas vectoriales de soporte.
3. Use como variable respuesta la variable categórica que especifica si la casa es barata, media o cara
4. 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).
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 que hizo.(donde el algoritmo se equivocó más, donde se equivocó menos y la importancia que tienen los errores).
8. Realice otro modelo de clasificación y compare sus resultados.
9. Genere un informe de los resultados y las explicaciones.

### EVALUACIÓN

- **(30 puntos).** Generación de varios modelos diferentes de SVM de los cambios en los parámetros.
- **(15 puntos).** Entrenamiento y predicción con los modelos generados.
- **(15 puntos).** Generación de las matrices de confusión y explicación de los resultados obtenidos
- **(20 puntos).** Comparación entre los modelos SVM
- **(20 puntos).** Comparación con la efectividad de otro algoritmo de clasificación.

#### **MATERIAL A ENTREGAR**

- Archivo .r o .py con el código y hallazgos comentados. Puede usar también Jupyter Notebooks
- Archivo PDF con las conclusiones y hallazgos encontrados.