Clasificación con máquina de vectores de soporte y random forest

Objetivos

Mediante esta actividad se pretende poner en práctica la creación de modelos basados en máquinas de vector de soporte y random forest. El objetivo es comprender de forma práctica las diferencias que existen a la hora de entrenar los diferentes modelos. De esta manera, lograrán:

* Comprender el tratamiento de datos que hay que realizar para los modelos de clasificación.
* Entender y aplicar el algoritmo de random forest a un problema de clasificación.
* Entender y aplicar el algoritmo de support vector machine a un problema de clasificación.
* Evaluar y analizar los resultados de los clasificadores.
* Investigar la aplicación de los modelos de clasificación a los problemas reales.

Pautas de elaboración

Debes realizar la actividad en el *notebook* adjunto y debes ir completar el código solicitado, así como responder a las preguntas que se van realizando.

* Análisis descriptivo de los datos. Debe completarse el código solicitado y responder a las preguntas. Todo ello en el *notebook* dado como base.
* Clasificación. Debe completarse el código solicitado y responder a las preguntas. Todo ello en el *notebook* dado como base.
* Investigación. Buscar un artículo científico en Google Académico que posea un caso de uso de regresión y que emplee una de las dos técnicas (o ambas) vistas en la actividad. Los artículos deben estar en revistas científicas y deben ser posteriores a 2015. No debe utilizar técnicas de *deep learning.*
* Para el artículo indicar:
  + El objetivo de la investigación; es decir, a qué problema real se está aplicando la clasificación.
  + La utilización de las técnicas de regresión y si realizan alguna adaptación de los algoritmos.
  + Los principales resultados de la aplicación y de la investigación.

Extensión y formato

Notebook Python.

Rúbrica

|  |  |  |  |
| --- | --- | --- | --- |
| Clasificación con máquina de vectores de soporte y *random forest* | Descripción | Puntuación máxima  (puntos) | Peso  % |
| Criterio 1 | Realiza el análisis descriptivo de los datos de manera adecuada | 2 | 20% |
| Criterio 2 | Realiza adecuadamente la aplicación del algoritmo de support vector machine | 3 | 30% |
| Criterio 3 | Realiza adecuadamente la aplicación del algoritmo de random forest | 3 | 30% |
| Criterio 4 | El artículo de investigación cumple con los criterios solicitados y analiza correctamente el trabajo | 2 | 20% |
|  |  | **10** | **100 %** |