

Programación de Inteligencia Artificial

Unidad 4: Aprendizaje supervisado y semisupervisado

Introducción

Durante el tema 4 hemos visto dos tipos distintos de aprendizaje: aprendizaje supervisado y aprendizaje semisupervisado.

Enunciado

Utilizando todo lo que ya sabes de los temas anteriores (**AED**, preprocesamiento de la información...) da solución a los dos problemas que se proponen a continuación.

Los problemas son **generalistas** a propósito (tienen muchas formas de ser resueltos dependiendo de las decisiones que vayas tomando). Son ejemplos perfectos de lo que podría solicitarse el día del examen final (de hecho, los del examen estarán más guiados, incluso). El primer problema está pensado para ser resuelto usando técnicas de aprendizaje supervisado y el segundo problema está pensado para ser resuelto usando técnicas de aprendizaje semisupervisado.

Aproximadamente, necesitarás unas 5 horas para resolver cada problema.

Problema 1: Sistema de arranque Tesla

Elon Musk se ha puesto en contacto con nuestra empresa para crear un sistema experto que determine, antes de arrancar el coche, la probabilidad de tener un accidente. Para ello, nos han otorgado acceso a un conjunto de datos pequeño, de 600 mediciones, en las que hubo (o no) accidentes.

- Realiza un **AED** sobre el conjunto de datos.
 - Estadísticos iniciales. **0.2 puntos**
 - Distribuciones de las variables numéricas del conjunto de datos. **0.3 puntos**
 - Matriz de correlación. **0.5 puntos**
- Realiza el preprocesamiento de datos de tu problema.
 - Reserva un conjunto de datos para validación y otro para testeo. **0.5 puntos**
 - Columnas inútiles, valores sin sentido y atípicos. **1 puntos**
 - Tratamiento de valores nulos. **0.5 puntos**
 - Análisis de variabilidad. **0.5 puntos**
 - Columnas categóricas. **0.5 punto**
 - Reducción de la dimensionalidad. **1 punto**
- Entrena y optimiza un modelo de KNN.
 - Entrenar un KNN. **0.4 puntos**
 - Optimizar el KNN con la técnica *GridSearch*. **0.6 puntos**
- Entrena y optimiza un modelo de DT.
 - Entrenar un DT. **0.3 puntos**
 - Explicar el DT. **0.1 puntos**
 - Optimizar el DT con la técnica *RandomSearch*. **0.5 puntos**
 - Explicar el DT. **0.1 puntos**
- Entrena y optimiza un modelo de SVM.
 - Entrenar un SVM. **0.7 puntos**
 - Optimizar el SVM con la técnica *GridSearch*. **0.3 puntos**
- Entrena y optimiza un modelo de NL.
 - Entrenar un NL. **0.4 puntos**
 - Optimizar el NL con la técnica *RandomSearch*. **0.6 puntos**
- Crear modelos *ensemble* usando:
 - Los tres mejores modelos obtenidos, usando el siguiente criterio: media aritmética de todos los modelos que tengan una fiabilidad superior al 80%. **0.7 puntos**
 - Todos los modelos, usando como criterio un modelo de Regresión Lineal. **0.3 puntos**

Problema 2: Fallos de producto

Ayuda a una empresa a mejorar su producto prediciendo cuándo fallará. El problema principal es que detectar, notificar y registrar los fallos es un proceso muy costoso, por lo que no se dispone de todas las etiquetas finales que permitirían entrenar perfectamente al modelo.

- Realiza un **AED** sobre el conjunto de datos.
 - Estadísticos iniciales. **0.2 puntos**
 - Distribuciones de las variables numéricas del conjunto de datos. **0.3 puntos**
 - Matriz de correlación. **0.5 puntos**
- Realiza el preprocesamiento de datos de tu problema.
 - Reserva un conjunto de datos para validación y otro para testeo. **0.5 puntos**
 - Columnas inútiles, valores sin sentido y atípicos. **1 puntos**
 - Tratamiento de valores nulos. **0.5 puntos**
 - Análisis de variabilidad. **0.5 puntos**
 - Columnas categóricas. **0.5 punto**
 - Reducción de la dimensionalidad. **1 punto**
- Realiza un etiquetado automático. **1 punto**
- Entrena y optimiza distintos modelos supervisados.
 - Modelo 1. **1 punto**
 - Modelo 2. **1 punto**
 - Modelo 3. **1 punto**
- Crea un modelo *ensemble* y explica el criterio que utilizas. **1 punto**

Entrega

Realiza tu tarea en dos documentos **ipynb** (en Jupyter o en Google Colab). Una vez finalizada, entrégala mediante el Aula Virtual en formato **zip** en la tarea reservada para ello.