

GUIONES DE PRÁCTICAS – MÓDULO 3

Práctica 2_2

Contexto

El mismo que la anterior.

Ejercicios

Sobre el RDD de instancias creado en la práctica anterior ve haciendo las siguientes operaciones:

1. Transforma el RDD de las instancias en un RDD numérico de forma que cada línea sea un array de valores numéricos. Para ello se ha de eliminar el último atributo que es la clase y de tipo nominal.
2. Calcula la suma de todos los valores de todos los atributos.
3. Calcula la suma de todos los valores de cada instancia (aunque esto no tendría sentido para el aprendizaje del modelo).
4. Comprueba que la suma calculada en el apartado 2 coincide con la que se haría sumando todos los resultados obtenidos en el apartado 3.
5. Calcula la suma de todos los valores para el atributo SepallLength, su media, valor máximo y mínimo. Se puede comprobar que estos dos últimos valores coinciden con los valores de rango indicados para el atributo en la cabecera.
6. Calcula lo mismo que en el anterior, la suma, media, valor máximo y mínimo pero para todos los atributos de entrada que hubiese.
7. Captura el DAG de alguna de las operaciones que has realizado.

Nota: no se deben utilizar las acciones max, min, mean, sum, sobre RDDs.

Entrega

Se debe entregar un fichero pdf en el que se especifique el enunciado de cada ejercicio, la solución en spark (con algún comentario explicativo) y captura de pantalla con la ejecución y los resultados obtenidos en la Shell.