Database Systems Homework Assignment 5

Name: Sai Venkata Naga Saketh Anne
USC ID: 3725520208
USC Email: annes@usc.edu

Part 1: Data Mining

Q1: Discuss the following Linear regression equation generated in the WEKA tool?

The Linear regression equation consists of 12 terms, i.e., it contains 11 fields from the given dataset that are multiplied with respective coefficients, and a numeric number that are required to calculate the MEDV (class) value. The Linear regression equation is as follows,

The Linear Regression Model to predict the MEDV(class) value is.

class =
-0.1084 * CRIM + 0.0458 * ZN +
2.7187 * CHAS=1 +
- 17.376 * NOX + 3.8016 * RM +
- 1.4927 * DIS +
0.2996 * AD + - 0.0118 * TAX +
- 0.9465 * PTRATIO + 0.0093 * B +
- 0.5226 * LSTAT + 36.3411

In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models.

As per the above linear regression equation, the MEDV(class) value depends on the following attributes, CRIM, ZN, CHAS, NOX, RM, DIS, AD, TAX, PTRATIO, B, LSTAT and a constant 36.3411.

Q2: What is the linear equation in the KNIME tool?

Number of rings = -0.8249 * sex=I + 0.0577 * sex=M + -0.4583 * length + 11.0751 * diameter + 10.7615 * height + 8.9754 * whole _weight + -19.7869 * shucked _weight + -10.5818 * viscera weight + 8.7418 * shell weight + 3.8946

The above is the linear regression equation generated by the KNIME tool to predict the number of rings based on all the given attributes(columns) in the given data.

Q3: What is the linear regression equation generated by the RapidMiner Studio tool?

Number of rings = -11.933 * length + 25.766 * diameter + 20.358 * height + 2.836

The above is the linear regression equation generated by the RapidMiner Studio tool to predict the number of rings based on the fields length, diameter, and height attributes(columns) in the given data.