YOUR QUESTIONS ANSWERED

Q: Will ROC and AUC will work in ordinal levels?

A) ROC and AUC are not necessarily appropriate assessment metrics for ordinal target variables. SAS Model Studio does not have built-in support for ordinal target variables, although you can choose to treat them as nominal or interval and then manually process the results.

Q: What is difference between WOE and target encoding?

A: The target encoding does what is called target likelihood encoding, and it works for both nominal and interval targets. Basically, for each categorical level, we calculate the overall mean of the target variable in that level (ideally using the training data so that we don't leak information from the validation data into the training process). Then we use that level-specific target mean as the numerical input value replacing the categorical value. Weight of evidence encoding (WOE) can only be used for binary or nominal targets (interval targets are not supported). WOE values are calculated based on whether the target value is an event or a non-event (thus the need for binary or nominal targets), and so don't just replace the categorical value with the target mean. This documentation page for PROC BINNING illustrates how WOE values are calculated in SAS. Note the WOE attribute expression in the documentation page to see the formulas for the non-adjusted and adjusted WOE values.

Q: What do each of the components in conn = swat.CAS( ) mean?

A: conn = swat.CAS(server-url = ..., server-port = ..., username = ..., password = ..., protocol = ...). The server-url is just the IP or URL of the SAS Viya server. The port is something that is configured when Viya is installed; it's the port that Viya is using to listen for connections from Python or R. The username and password can be replaced with an .authinfo file or with token authentication. The protocol is the method used to pass information between Python and Viya; this can be binary or http (REST using JSON). For more details, check this documentation page.

Q: If we do not store as a dataframe, does anything change in the commands for creating the plots after that?

A: If the data is not stored in a dataframe, we cannot plot it locally using Python commands. The CAS server (the part of SAS Viya that does the distributed computing) doesn't have any built-in plotting routines, so

all of the plotting in Viya is done by the graphical web applications. If you want to plot programmatically using Python code, you must bring the data locally in the form of a dataframe. Of course, once the data is in Python, you can use any Python plotting package you want.