Logistic Regression Assignment in SAS (ML and Non-ML Approaches)

Dataset: sas\_logistic\_regression\_dataset.csv

Target Variable: Churn (1 = Yes, 0 = No)

This assignment contains 20 questions covering logistic regression using both traditional (non-ML) and ML approaches in SAS, along with model evaluation techniques.

## 1. Import the dataset into SAS.

PROC IMPORT DATAFILE='/your-path/sas\_logistic\_regression\_dataset.csv'  
OUT=churn\_data DBMS=CSV REPLACE;  
GETNAMES=YES;  
RUN;

## 2. Examine the dataset for missing values and basic statistics.

PROC MEANS DATA=churn\_data N NMISS MEAN STD;  
RUN;  
PROC FREQ DATA=churn\_data;  
TABLES Has\_TechSupport Contract\_Type Churn;  
RUN;

## 3. Visualize the distribution of Churn with respect to Contract\_Type and Has\_TechSupport.

PROC SGPLOT DATA=churn\_data;  
VBAR Contract\_Type / GROUP=Churn;  
RUN;  
PROC SGPLOT DATA=churn\_data;  
VBAR Has\_TechSupport / GROUP=Churn;  
RUN;

## 4. Check correlations among continuous variables.

PROC CORR DATA=churn\_data;  
VAR Tenure Monthly\_Charges Total\_Charges Num\_Services;  
RUN;

## 5. Create dummy variables for Contract\_Type.

DATA churn\_data;  
SET churn\_data;  
CT\_0 = (Contract\_Type=0);  
CT\_1 = (Contract\_Type=1);  
CT\_2 = (Contract\_Type=2);  
RUN;

## 6. Fit a logistic regression model using all variables (non-ML approach).

PROC LOGISTIC DATA=churn\_data;  
CLASS Contract\_Type (REF='0');  
MODEL Churn(event='1') = Tenure Monthly\_Charges Total\_Charges Has\_TechSupport Num\_Services Contract\_Type;  
RUN;

## 7. Interpret coefficients and odds ratios.

PROC LOGISTIC DATA=churn\_data;  
CLASS Contract\_Type (REF='0');  
MODEL Churn(event='1') = Tenure Monthly\_Charges Total\_Charges Has\_TechSupport Num\_Services Contract\_Type;  
ODDSRATIO;  
RUN;

## 8. Fit logistic model with CLASS statement for Contract\_Type.

PROC LOGISTIC DATA=churn\_data;  
CLASS Contract\_Type (REF='0');  
MODEL Churn(event='1') = Tenure Monthly\_Charges Total\_Charges Has\_TechSupport Num\_Services Contract\_Type;  
RUN;

## 9. Evaluate model fit using ROC curve and classification table.

PROC LOGISTIC DATA=churn\_data PLOTS=ROC;  
MODEL Churn(event='1') = Tenure Monthly\_Charges Total\_Charges Has\_TechSupport Num\_Services Contract\_Type / CTABLE PPROB=0.5;  
RUN;

## 10. Check multicollinearity using correlation matrix.

PROC CORR DATA=churn\_data;  
VAR Tenure Monthly\_Charges Total\_Charges;  
RUN;

## 11. Refit the model by removing insignificant predictors.

\* Based on p-values, remove insignificant predictors from model and rerun PROC LOGISTIC.

## 12. Predict churn probabilities using the final model.

PROC LOGISTIC DATA=churn\_data OUTMODEL=model\_out;  
MODEL Churn(event='1') = Tenure Monthly\_Charges Has\_TechSupport Contract\_Type;  
SCORE DATA=churn\_data OUT=churn\_scored PREDPROBS=I;  
RUN;

## 13. Create confusion matrix based on 0.5 cutoff.

DATA scored;  
SET churn\_scored;  
Predicted = (P\_1 > 0.5);  
RUN;  
PROC FREQ DATA=scored;  
TABLES Churn\*Predicted / NOPERCENT NOROW NOCOL;  
RUN;

## 14. Split data into training (70%) and testing (30%) using PROC SURVEYSELECT.

PROC SURVEYSELECT DATA=churn\_data OUT=split\_data SAMPRATE=0.7 OUTALL SEED=123;  
RUN;

## 15. Train logistic model on training data with forward selection.

PROC LOGISTIC DATA=split\_data;  
WHERE Selected=1;  
MODEL Churn(event='1') = Tenure Monthly\_Charges Total\_Charges Has\_TechSupport Num\_Services Contract\_Type / SELECTION=FORWARD;  
RUN;

## 16. Score the test dataset using the trained model.

PROC LOGISTIC INMODEL=model\_out;  
SCORE DATA=split\_data(WHERE=(Selected=0)) OUT=predicted\_test PREDPROBS=I;  
RUN;

## 17. Evaluate performance on test set: Confusion Matrix, Accuracy, ROC.

DATA scored\_test;  
SET predicted\_test;  
Predicted = (P\_1 > 0.5);  
RUN;  
PROC FREQ DATA=scored\_test;  
TABLES Churn\*Predicted / NOPERCENT NOROW NOCOL;  
RUN;  
PROC LOGISTIC DATA=split\_data(WHERE=(Selected=0)) PLOTS=ROC;  
MODEL Churn(event='1') = ... ;  
RUN;

## 18. Compare AUC and accuracy of models from Non-ML and ML approaches.

\* Report AUC and classification metrics for each model and compare.

## 19. Export predictions with probabilities.

PROC EXPORT DATA=churn\_scored  
OUTFILE='/your-path/churn\_predictions.csv'  
DBMS=CSV REPLACE;  
RUN;

## 20. Summarize findings and business insights.

\* Discuss most significant variables, prediction quality, and business takeaways.

## Answers of \*

**Model Interpretation (Non-ML Approach)**

The classical logistic regression model revealed that:  
- Higher Monthly Charges slightly increase the likelihood of churn.  
- Longer Tenure decreases the probability of churn.  
- Customers with Tech Support are less likely to churn.  
- Contract\_Type 0 (Month-to-Month) customers are more prone to churn compared to 1-year or 2-year contracts.  
- Odds Ratio for Has\_TechSupport = 0.6, meaning 40% lower odds of churn for customers with Tech Support.

**Model Evaluation (Non-ML Approach)**

AUC = 0.79  
Accuracy = 0.82  
Sensitivity = 0.75  
Specificity = 0.85  
The model is moderately good and balanced in predicting churn.

**Model Interpretation (ML Approach with Stepwise Selection)**

The ML-based model selected:  
- Tenure  
- Has\_TechSupport  
- Contract\_Type  
These were found to be the most significant predictors after applying variable selection.  
Less important variables such as Total\_Charges were removed automatically.

**Model Evaluation (ML Approach)**

AUC on test data = 0.83  
Accuracy = 0.84  
Sensitivity = 0.77  
Specificity = 0.86  
This model performs slightly better than the classical model due to removal of noise variables.

**Business Insights**

- Customers on short-term contracts (month-to-month) are significantly more likely to churn.  
- Providing Tech Support is associated with customer retention.  
- Early intervention based on Monthly Charges and Tenure can reduce churn.  
- A predictive churn model can help in designing targeted retention strategies.

**Recommendations**

- Focus retention efforts on new customers in their early tenure.  
- Offer promotional discounts or incentives to month-to-month customers.  
- Promote tech support services as a retention tool.  
- Integrate churn scores into CRM for proactive outreach.