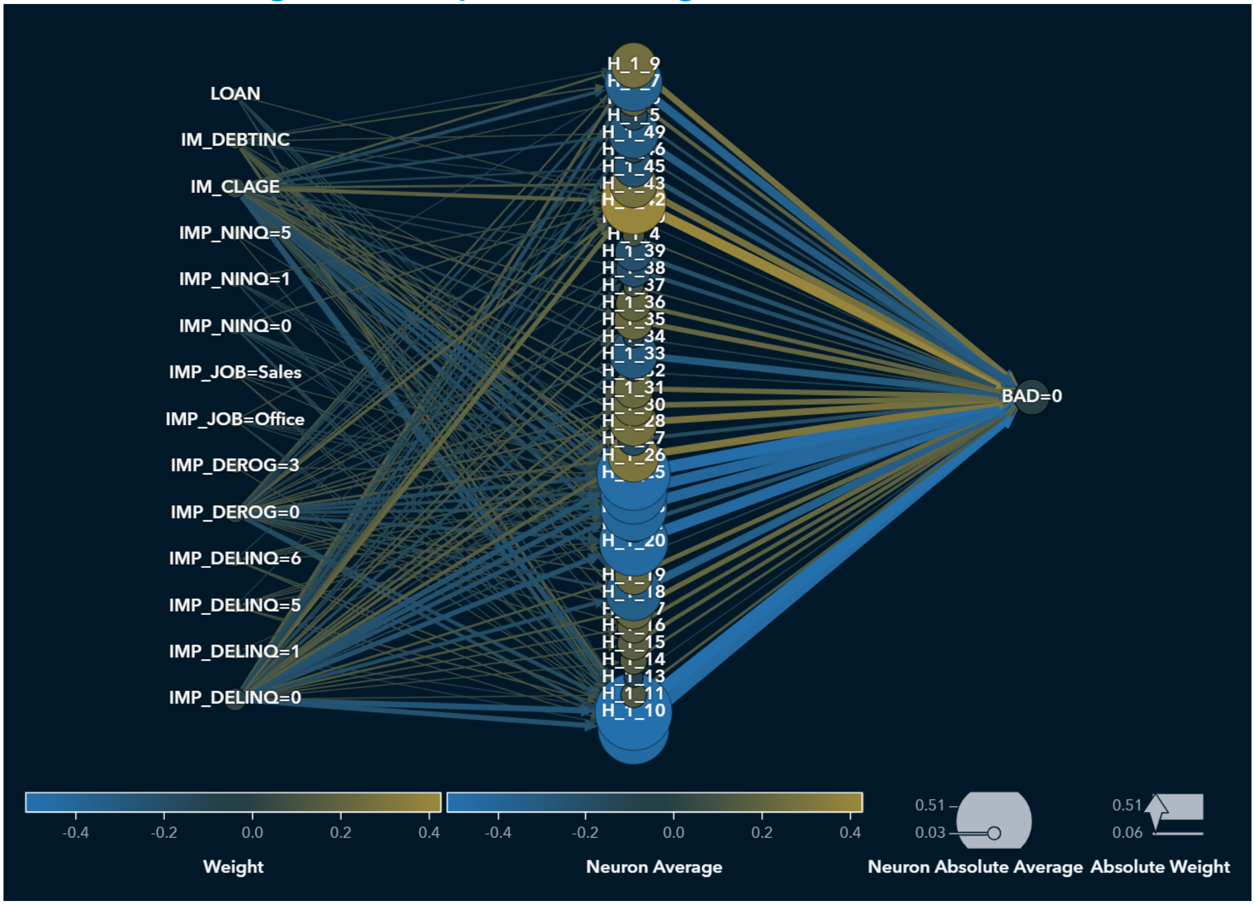
Machine Learning Analytic

"Neural Network" Results

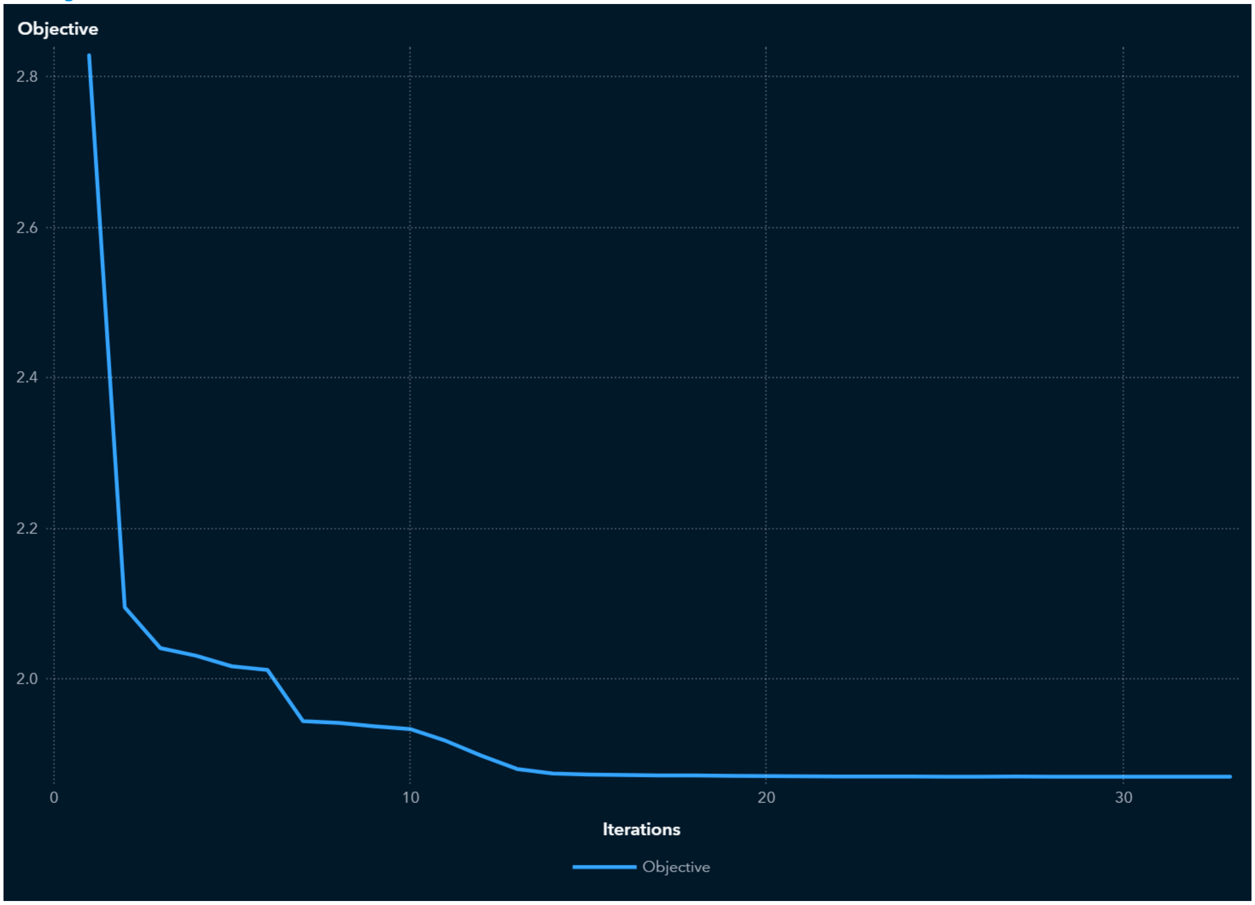
by: jbae7@ncsu.edu

Contents

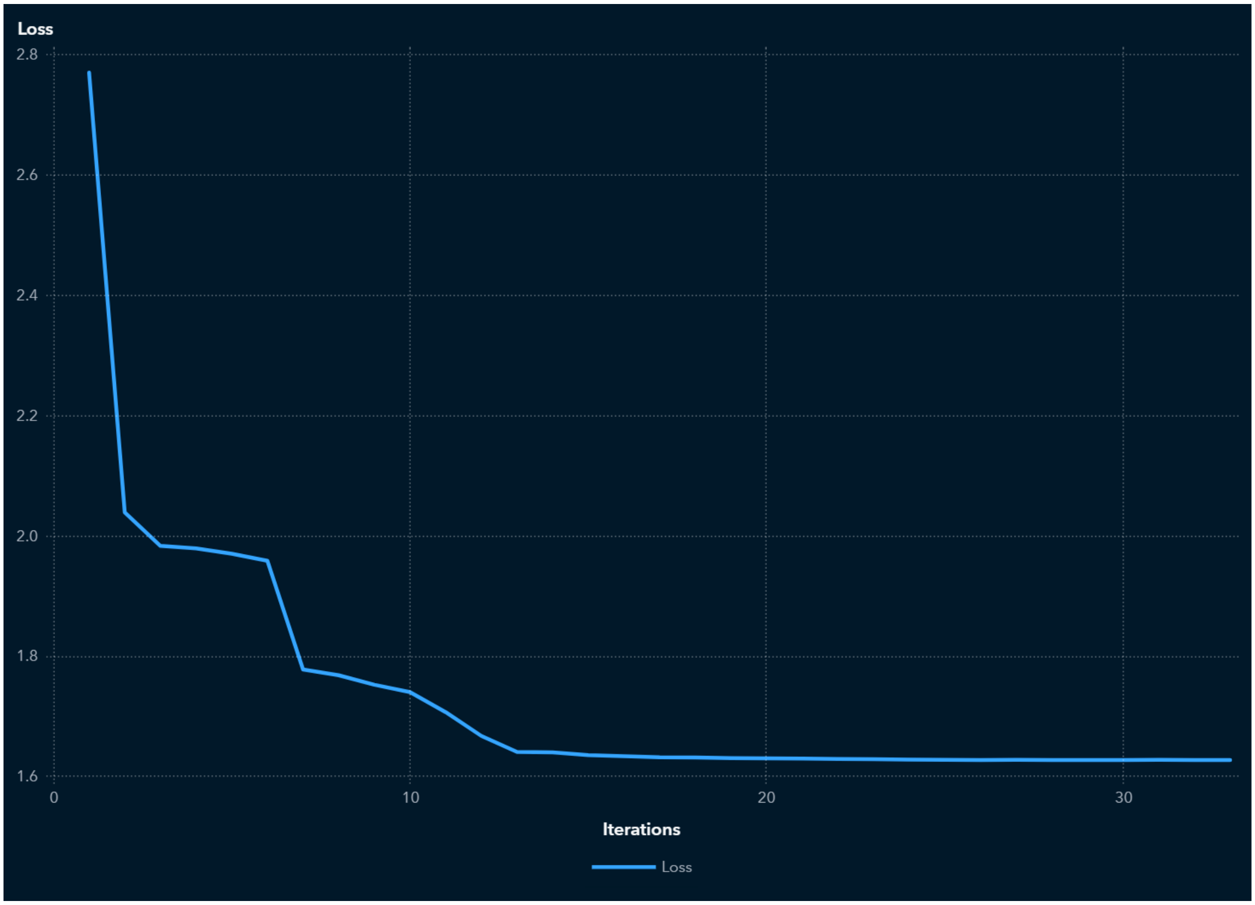
Network Diagram: Top 200 Weights



Objective



Loss



Score Inputs

Name	Role	Variable Level	Type
DELINQ	INPUT	NOMINAL	N
DEROG	INPUT	NOMINAL	N
IM_CLAGE	INPUT	INTERVAL	N
IM_CLNO	INPUT	INTERVAL	N
IM_DEBTINC	INPUT	INTERVAL	N
JOB	INPUT	NOMINAL	C
LOAN	INPUT	INTERVAL	N
NINQ	INPUT	NOMINAL	N
REASON	INPUT	BINARY	C
VALUE	INPUT	INTERVAL	N

Variable Type	Variable Label	Variable Format	Variable Length
double			8
double			8
double			8
double			8
double			8
char			7
double			8
double			8
char			7
double			8

Score Outputs

Name	Role	Type	Variable Type
EM_CLASSIFICATION	CLASSIFICATION	C	char
EM_EVENTPROBABILITY	PREDICT	N	double
EM_PROBABILITY	PREDICT	N	double
IMP_DELINQ	INPUT	N	double
IMP_DEROG	INPUT	N	double
IMP_JOB	INPUT	C	char
IMP_NINQ	INPUT	N	double
IMP_REASON	INPUT	C	char
IMP_VALUE	REJECTED	N	double
I_BAD	CLASSIFICATION	C	char
P_BAD0	PREDICT	N	double
P_BAD1	PREDICT	N	double

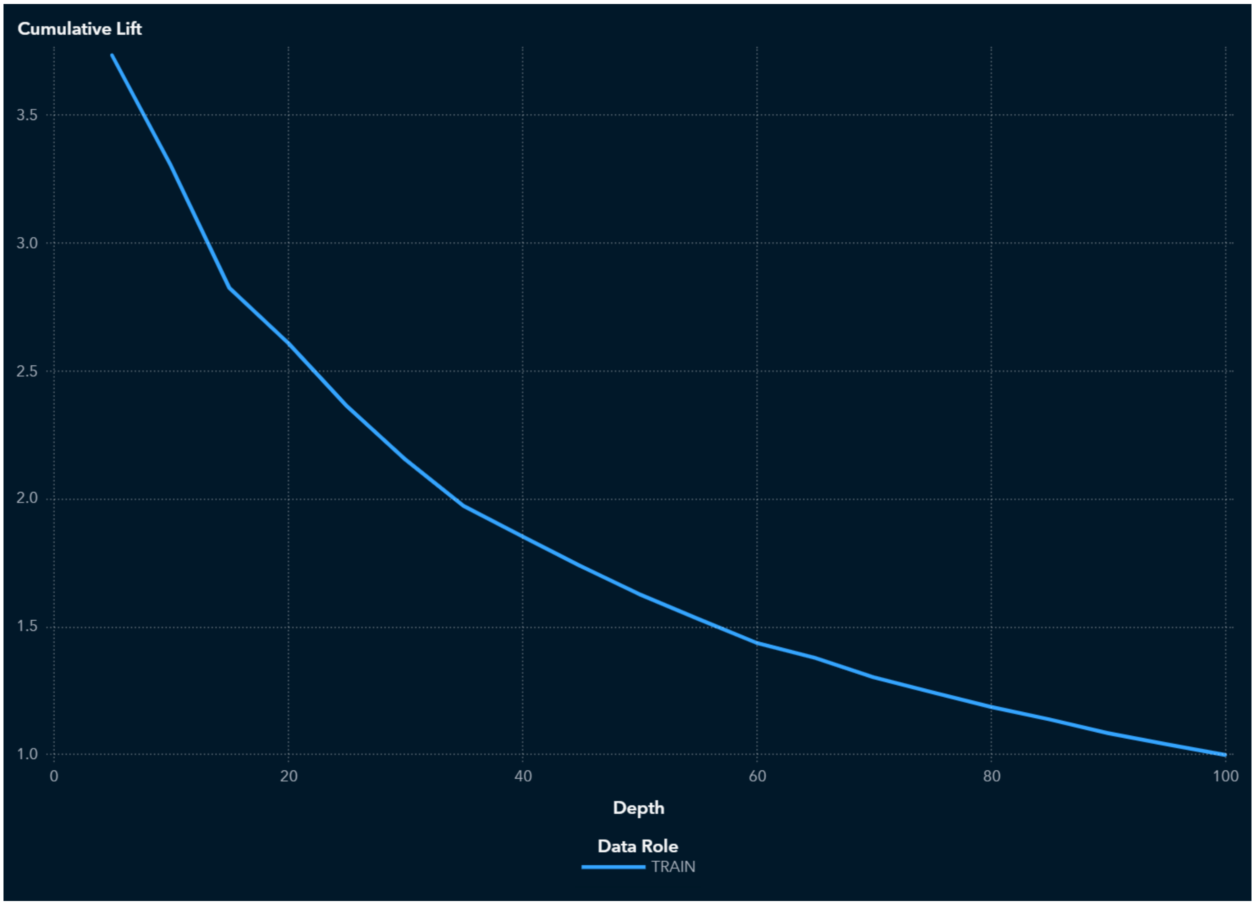
Variable Label	Variable Format	Variable Length	Creator
Predicted for BAD		12	neural
Probability for BAD=1		8	neural
Probability of Classification		8	neural
Imputed DELINQ		8	impute
Imputed DEROG		8	impute
Imputed JOB		7	impute
Imputed NINQ		8	impute
Imputed REASON		7	impute
Imputed VALUE		8	impute
Into: BAD		12	neural
Predicted: BAD=0		8	neural

Variable Label	Variable Format	Variable Length	Creator
Predicted: BAD=1		8	neural

Function	Creator GUID
CLASSIFICATION	a62d8b98-b547-4190-b92f-4db805d4ff76
PREDICT	a62d8b98-b547-4190-b92f-4db805d4ff76
PREDICT	a62d8b98-b547-4190-b92f-4db805d4ff76
TRANSFORM	b727c11b-1371-4c1d-8746-51b94c237069
TRANSFORM	b727c11b-1371-4c1d-8746-51b94c237069
TRANSFORM	b727c11b-1371-4c1d-8746-51b94c237069
TRANSFORM	b727c11b-1371-4c1d-8746-51b94c237069
TRANSFORM	b727c11b-1371-4c1d-8746-51b94c237069
TRANSFORM	b727c11b-1371-4c1d-8746-51b94c237069
CLASSIFICATION	a62d8b98-b547-4190-b92f-4db805d4ff76
PREDICT	a62d8b98-b547-4190-

Function	Creator GUID
	b92f-4db805d4ff76
PREDICT	a62d8b98- b547-4190- b92f-4db805d4ff76

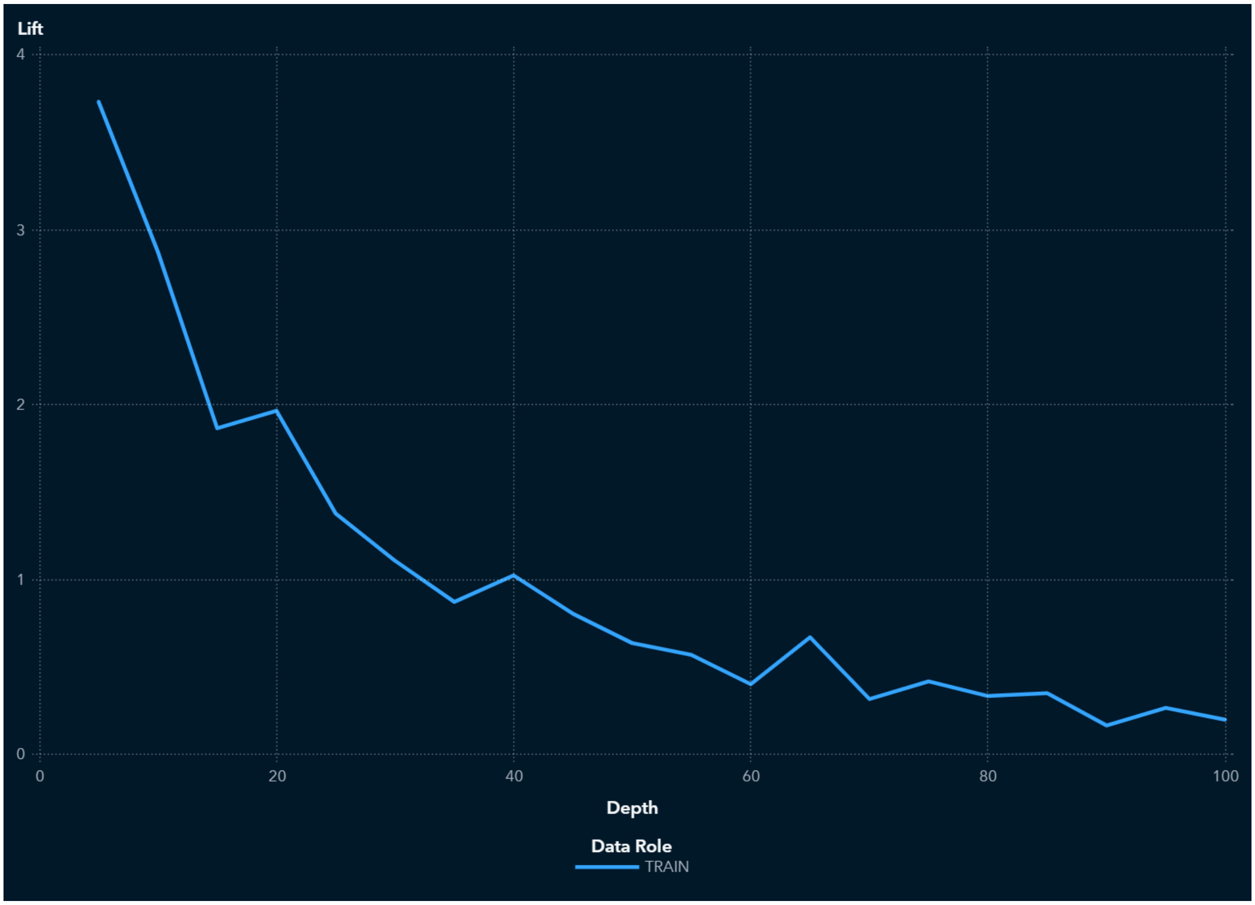
Cumulative Lift



The TRAIN partition has a Cumulative Lift of 3.31 in the 10% quantile (depth of 10) meaning there are 3.31 times more events in the first two quantiles than expected by random (10% of the total number of events). Because this value is greater than 1, it is better to use your model to identify responders than no model, based on the selected partition.

Cumulative lift is calculated by sorting each partition in descending order by the predicted probability of the target event P_BAD1, which represents the predicted probability of the event "1" for the target BAD. The data is divided into 20 quantiles (demi-deciles, with 5% of the data in each), and the number of events in each quantile is computed. The cumulative lift for a particular quantile is the ratio of the number of events across all quantiles up to and including the current quantile to the number of events that would be there at random, or equivalently, the ratio of the cumulative response percentage to the baseline response percentage. The cumulative lift at depth 10 includes the top 10% of the data, which is the first 2 quantiles, which would have 10% of the events at random. Thus, cumulative lift measures how much more likely it is to observe an event in the quantiles than by selecting observations at random.

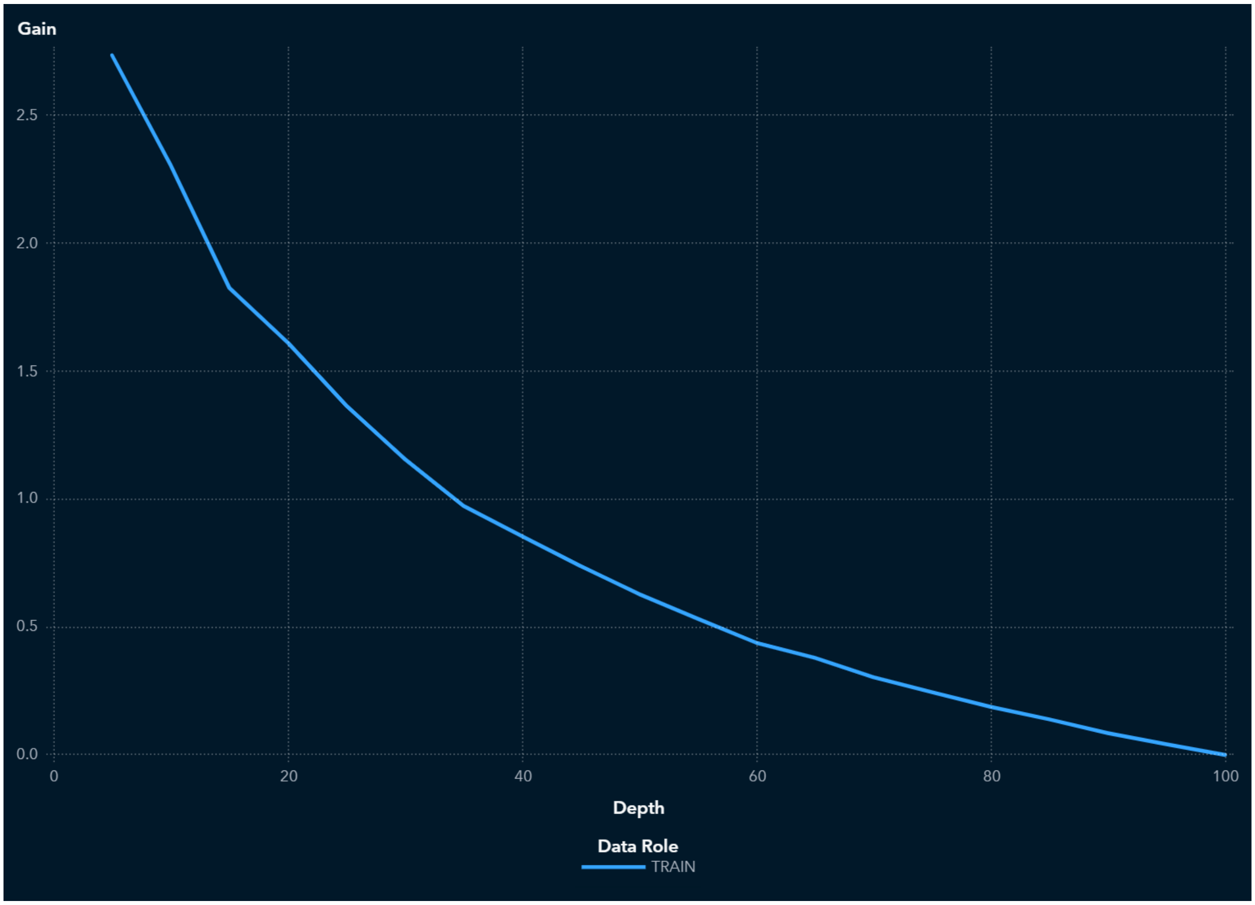
Lift



The TRAIN partition has a Lift of 3.73 in the 5% quantile (depth of 5) meaning there are 3.73 times more events in that quantile than expected by random (5% of the total number of events). Because this value is greater than 1, it is better to use your model to identify responders than no model, based on the selected partition.

Lift is calculated by sorting each partition in descending order by the predicted probability of the target event P_BAD1, which represents the predicted probability of the event "1" for the target BAD. The data is divided into 20 quantiles (demi-deciles, with 5% of the data in each), and the number of events in each quantile is computed. Lift is the ratio of the number of events in that quantile to the number of events that would be there at random, or equivalently, the ratio of the response percentage to the baseline response percentage. With 20 quantiles, it is expected that 5% of the events occur in each quantile. Thus, Lift measures how much more likely it is to observe an event in each quantile than by selecting observations at random.

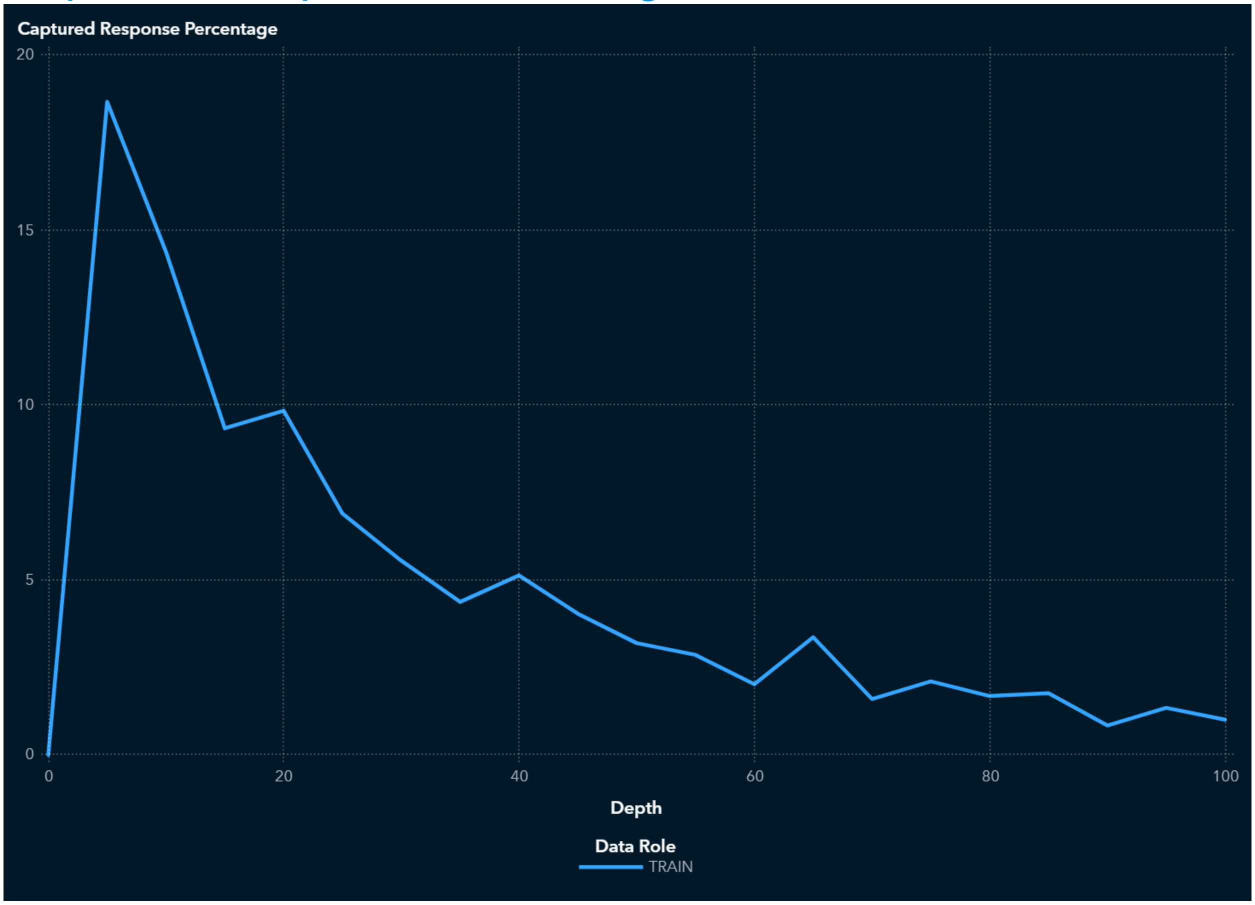
Gain



The TRAIN partition has a Gain of 2.3 at the 10% quantile (depth of 10). Because this value is greater than 0, it is better to use your model to identify responders than no model, based on the selected partition. The best possible value of Gain for this partition at depth 10 is 4.01.

Gain is calculated by sorting each partition in descending order by the predicted probability of the target event P_BAD1, which represents the predicted probability of the event "1" for the target BAD. The data is divided into 20 quantiles (demi-deciles, with 5% of the data in each), and the number of events in each quantile is computed. Gain is a cumulative measure for the quantiles up to an including the current one and is calculated as (number of events in the quantiles) / (number of events expected by random) - 1. With 20 quantiles, it is expected that 5% of the events occur in each quantile. Note that the value of Gain is the same as the value of Cumulative Lift - 1. If the value of Gain is greater than 0, then your model is better at identifying events than using no model.

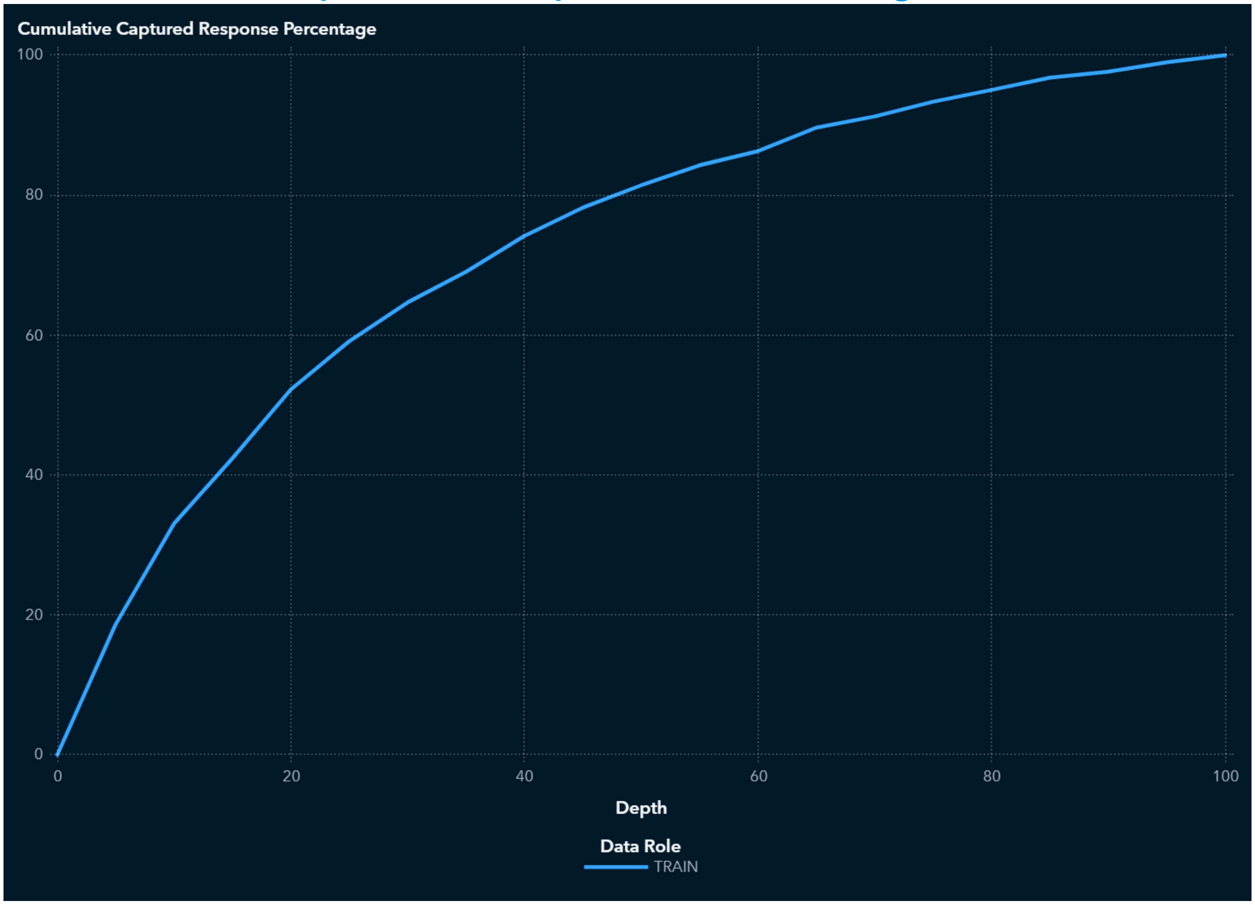
Captured Response Percentage



At the 5% quantile (depth of 5), the TRAIN partition has a Captured response percentage of 18.7 (compared to the expected value of 5 for no model). The best possible value of Captured response percentage for this partition at depth 5 is 25.06.

Captured response percentage is calculated by sorting each partition in descending order by the predicted probability of the target event P_BAD1, which represents the predicted probability of the event "1" for the target BAD. The data is divided into 20 quantiles (demi-deciles, with 5% of the data in each), and the number of events in each quantile is computed. Captured response percentage is the percentage of the total number of events that are in that quantile. With no model, it is expected that 5% of the events are in each quantile.

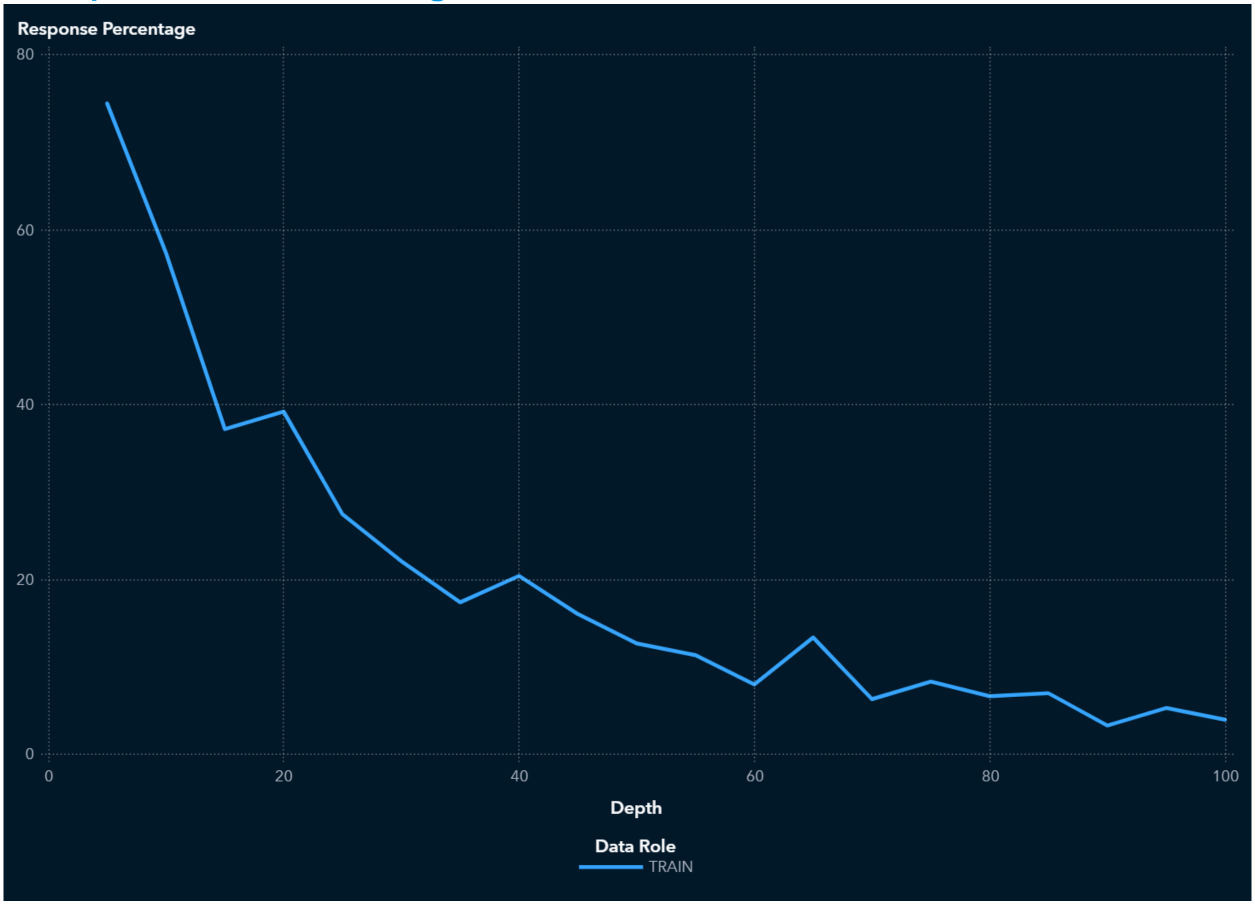
Cumulative Captured Response Percentage



In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative captured response percentage of 33.1 (compared to the expected value of 10 for no model). The best possible value of Cumulative captured response percentage for this partition at depth 10 is 50.13.

Cumulative captured response percentage is calculated by sorting each partition in descending order by the predicted probability of the target event P_BAD1, which represents the predicted probability of the event "1" for the target BAD. The data is divided into 20 quantiles (demi-deciles, with 5% of the data in each), and the number of events in each quantile is computed. The cumulative captured response percentage for a particular quantile is the percentage of the total number of events that are in the quantiles up to and including the current quantile. With no model, it is expected that 5% of the events are in each quantile, so the cumulative captured response percentage at depth 10 would be 10%.

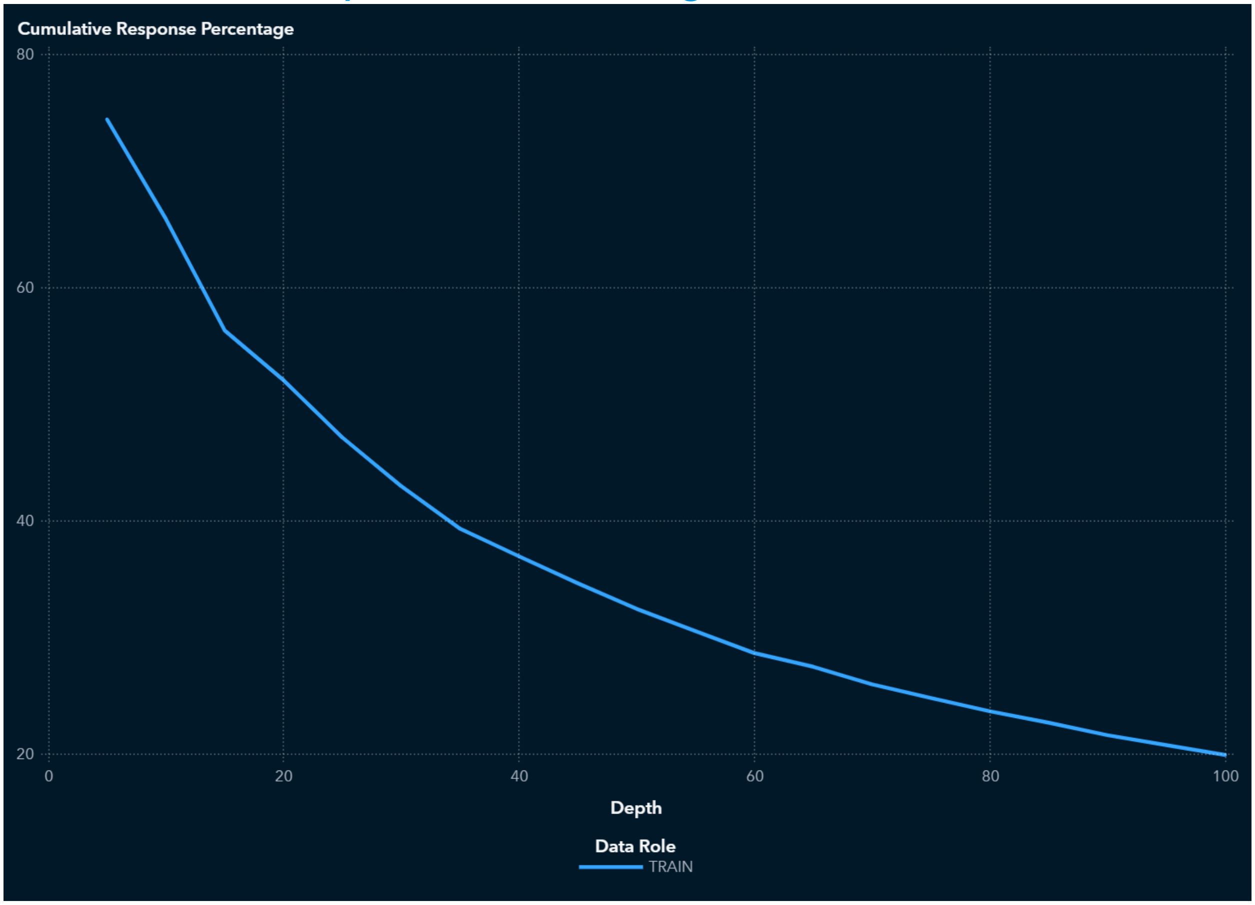
Response Percentage



At the 5% quantile (depth of 5), the TRAIN partition has a Response percentage of 74.5. The best possible value of Response percentage for this partition at depth 5 is 100.

Response percentage is calculated by sorting each partition in descending order by the predicted probability of the target event P_BAD1, which represents the predicted probability of the event "1" for the target BAD. The data is divided into 20 quantiles (demi-deciles, with 5% of the data in each), and the number of events in each quantile is computed. Response percentage is the percentage of observations that are events in that quantile. With no model, it is expected that the response percentage is constant across quantiles, 100*overall-event-rate. This is also called the baseline response percentage.

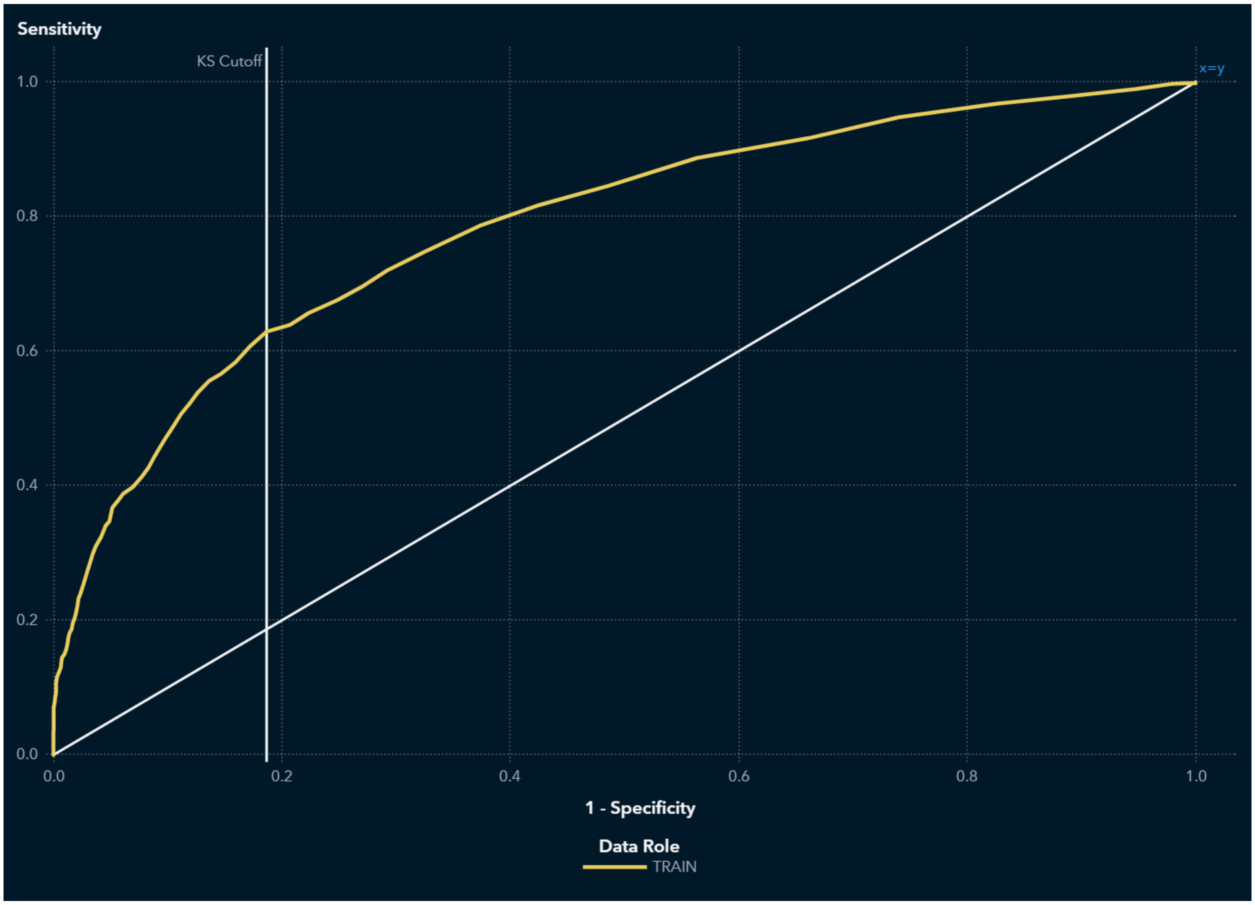
Cumulative Response Percentage



In the top 10% of the data (depth 10), the TRAIN partition has a Cumulative response percentage of 65.9. The best possible value of Cumulative response percentage for this partition at depth 10 is 100.

Cumulative response percentage is calculated by sorting in descending order each partition of the data by the predicted probability of the target event P_BAD1, which represents the predicted probability of the event "1" for the target BAD. The data is divided into 20 quantiles (demi-deciles, with 5% of the data in each), and the number of events in each quantile is computed. The cumulative response percentage for a particular quantile is the percentage of observations that are events in the quantiles up to and including the current quantile. With no model, it is expected that the response percentage is constant across quantiles, 100*overall-event-rate. This is also called the baseline response percentage.

ROC



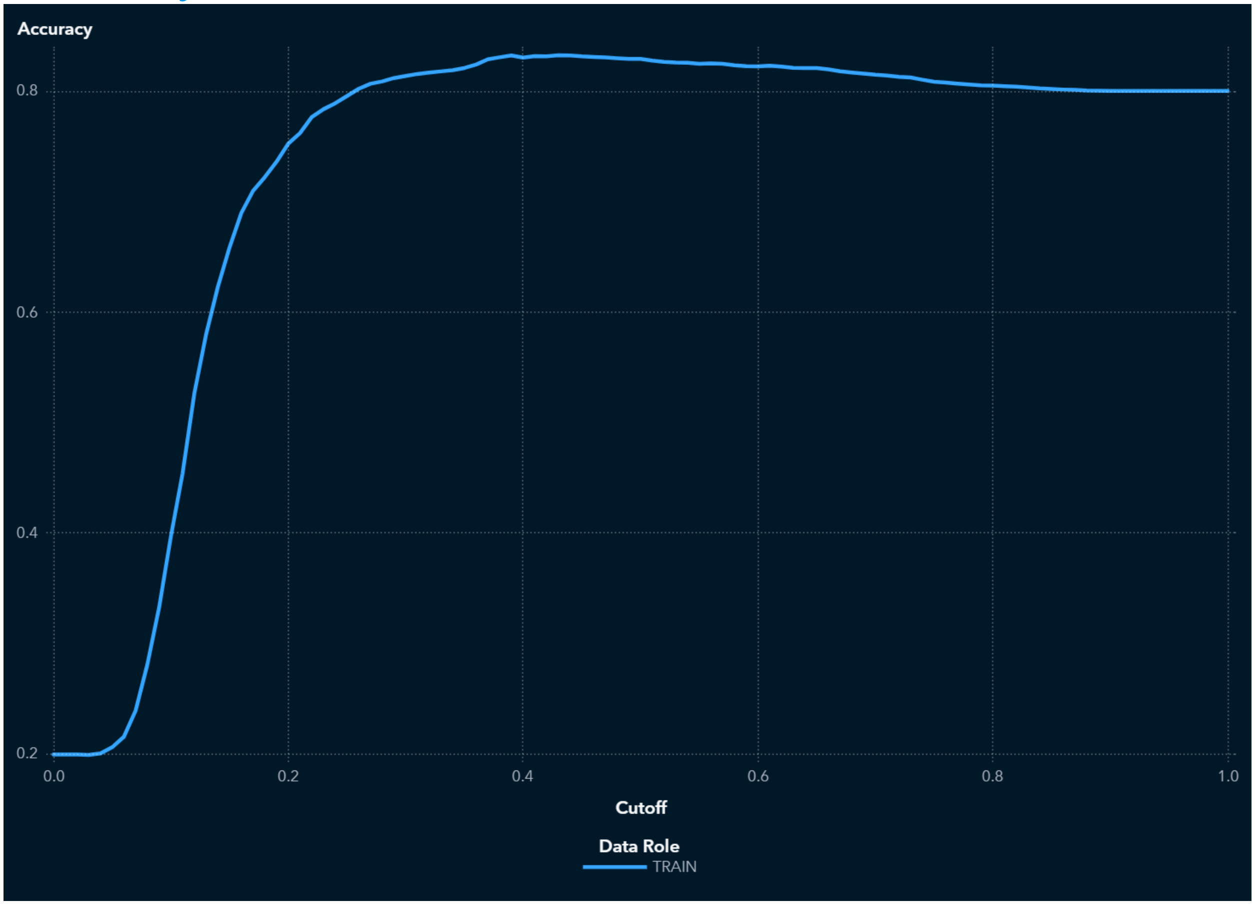
The ROC curve is a plot of sensitivity (the true positive rate) against 1-specificity (the false positive rate), which are both measures of classification based on the confusion matrix. These measures are calculated at various cutoff values. To help identify the best cutoff to use when scoring your data, the KS Cutoff reference line is drawn at the value of 1-specificity where the greatest difference between sensitivity and 1-specificity is observed for the TRAIN partition. The KS Cutoff line is drawn at the cutoff value 0.22, where the 1-specificity value is 0.187 and the sensitivity value is 0.629.

Cutoff values range from 0 to 1, inclusive, in increments of 0.01. At each cutoff value, the predicted target classification is determined by whether P_BAD1, which is the predicted probability of the event "1" for the target BAD, is greater than or equal to the cutoff value. When P_BAD1 is greater than or equal to the cutoff value, then the predicted classification is the event, otherwise it is a non-event.

The confusion matrix for each cutoff value contains four cells that display the true positives for events that are correctly classified (TP), false positives for non-events that are classified as events (FP), false negatives for events that are classified as non-events (FN), and true negatives for non-events that are classified as non-events (TN). True negatives include non-event classifications that specify a different non-event. Sensitivity is calculated as TP / (TP + FN). Specificity, the true negative rate, is calculated as TN / (TN + FP), so 1-specificity is FP / (TN + FP). The values of sensitivity and 1-specificity are plotted at each cutoff value.

A ROC curve that rapidly approaches the upper-left corner of the graph, where the difference between sensitivity and 1-specificity is the greatest, indicates a more accurate model. A diagonal line where sensitivity = 1-specificity indicates a random model.

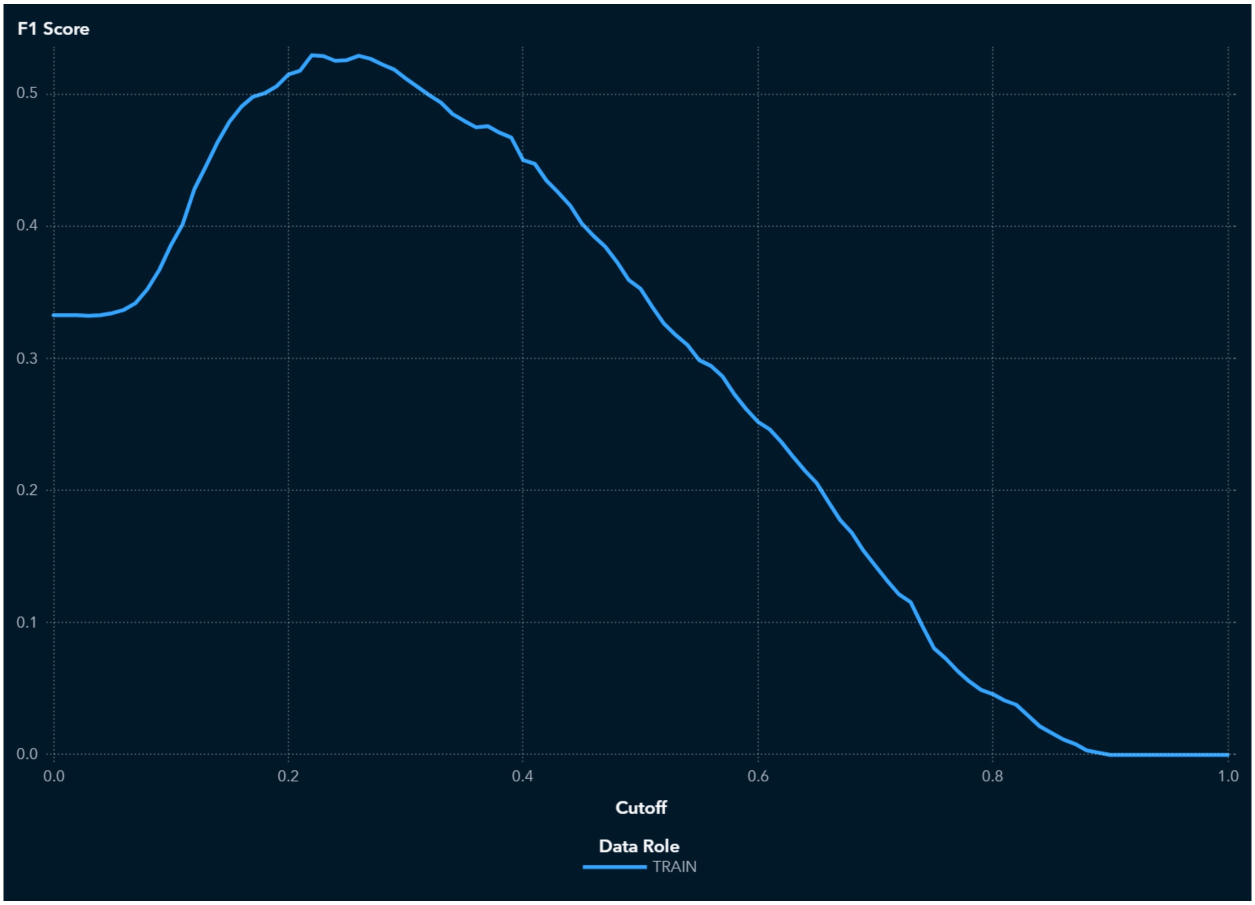
Accuracy



For this model, the accuracy in the TRAIN partition at the cutoff of 0.5 is 0.829.

Accuracy is the proportion of observations that are correctly classified as either an event or non-event, calculated at various cutoff values. Cutoff values range from 0 to 1, inclusive, in increments of 0.01. At each cutoff value, the predicted target classification is determined by whether P_BAD1, which is the predicted probability of the event "1" for the target BAD, is greater than or equal to the cutoff value. When P_BAD1 is greater than or equal to the cutoff value, then the predicted classification is the event, otherwise it is a non-event. When the predicted classification and the actual classification are both events (true positives) or both non-events (true negatives), the observation is correctly classified. If the predicted classification and actual classification disagree, then the observation is incorrectly classified. Accuracy is calculated as (true positives + true negatives) / (total observations).

F1 Score



For this model, the F1 score in the TRAIN partition at the cutoff of 0.5 is 0.353.

The F1 score combines the measures of precision and recall (or sensitivity), which are measures of classification based on the confusion matrix that are calculated at various cutoff values. Cutoff values range from 0 to 1, inclusive, in increments of 0.01. At each cutoff value, the predicted target classification is determined by whether P_BAD1, which is the predicted probability of the event "1" for the target BAD, is greater than or equal to the cutoff value. When P_BAD1 is greater than or equal to the cutoff value, then the predicted classification is the event, otherwise it is a non-event.

The confusion matrix for each cutoff value contains four cells that display the true positives for events that are correctly classified (TP), false positives for non-events that are classified as events (FP), false negatives for events that are classified as non-events (FN), and true negatives for non-events that are classified as non-events (TN). True negatives include non-event classifications that specify a different non-event.

Precision is calculated as TP / (TP + FP), and recall (or sensitivity) is calculated as TP /

(TP + FN). The F1 score is calculated as 2*Precision*Recall / (Precision + Recall), which is the harmonic mean of Precision and Recall. Larger F1 scores indicate a more accurate model.

Fit Statistics

Target Name	Data Role	Number of Observations	Average Squared Error
BAD	TRAIN	5,960	0.1260

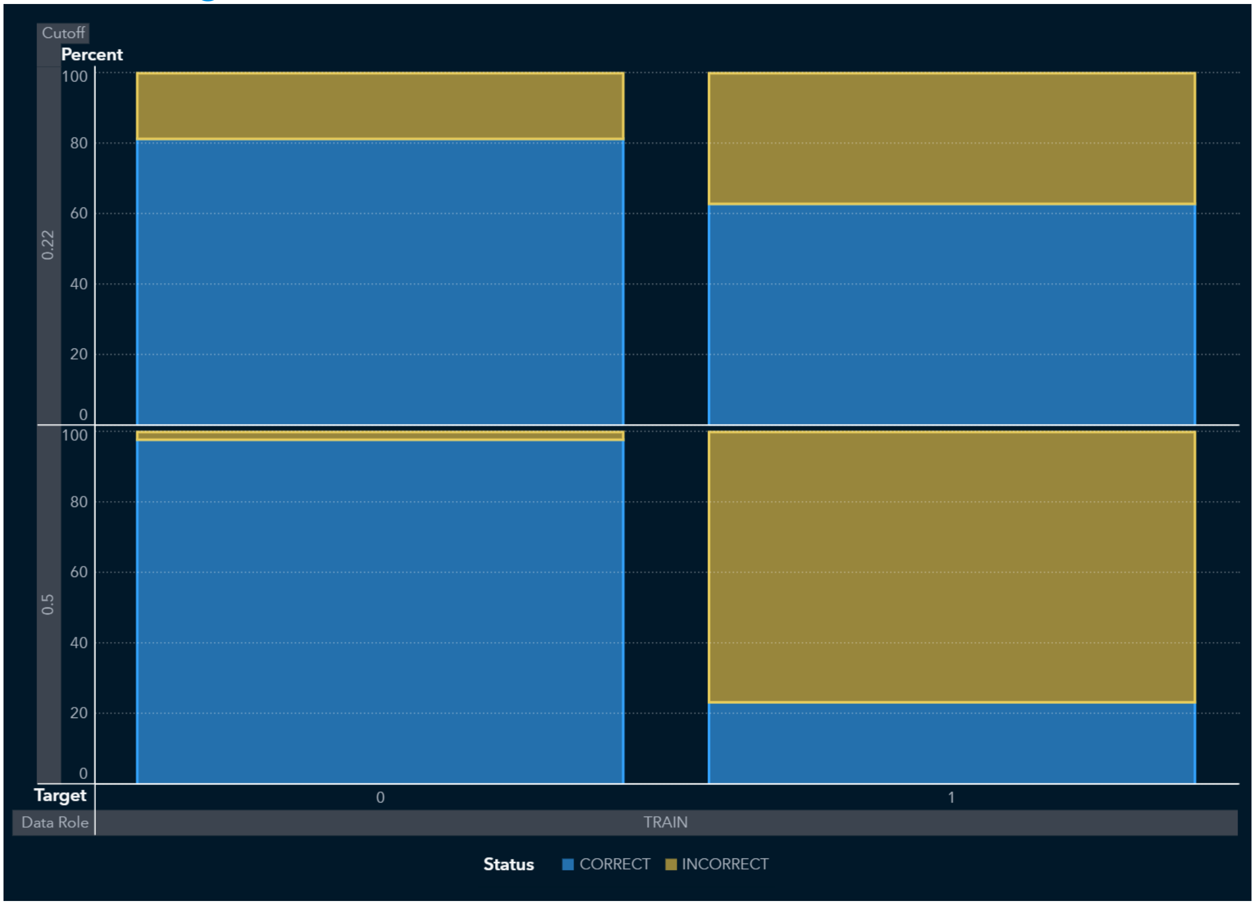
Divisor for ASE	Root Average Squared Error	Misclassification Rate	Multi-Class Log Loss
5,960	0.3549	0.1706	0.4070

KS (Youden)	Area Under ROC	Gini Coefficient	Gamma
0.4426	0.7874	0.5749	0.5882

Tau	KS Cutoff	KS at User-Specified Cutoff	Misclassification Rate at KS Cutoff (Event)
0.1836	0.2200	0.2110	0.2233

Misclassification Rate (Event)
0.1706

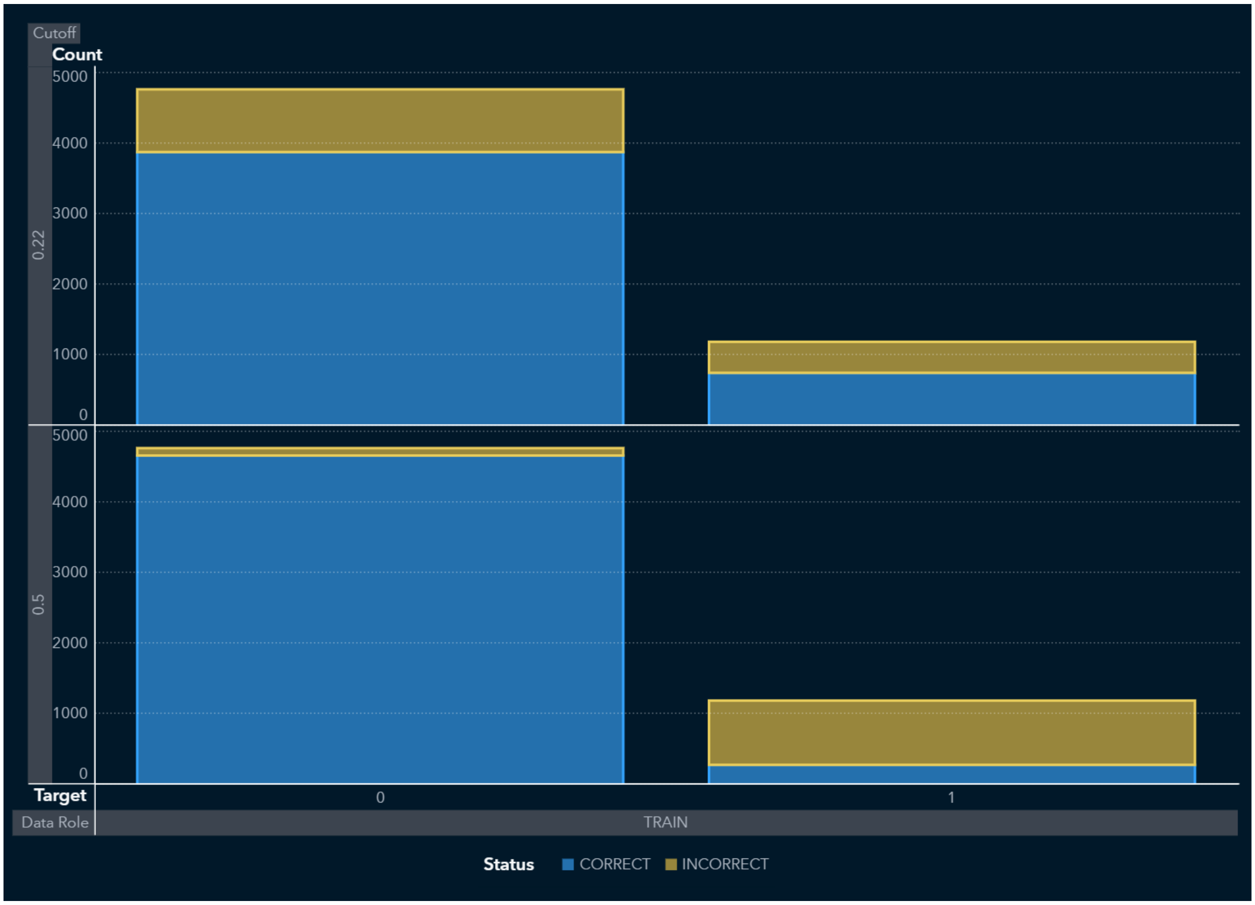
Percentage Plot



The Event Classification report is a visual representation of the confusion matrix at various cutoff values for each partition. The classification cutoffs used in the plot are the default (0.5) and the KS cutoff value 0.22 for the TRAIN partition.

For this data, for the bar corresponding to the event level of BAD, "   1", the segment of the bar colored as "CORRECT" corresponds to true positives.

Count Plot



The Event Classification report is a visual representation of the confusion matrix at various cutoff values for each partition. The classification cutoffs used in the plot are the default (0.5) and the KS cutoff value 0.22 for the TRAIN partition.

For this data, for the bar corresponding to the event level of BAD, "   1", the segment of the bar colored as "CORRECT" corresponds to true positives.

Table

Cutoff	Cutoff Source	Target Name	Response
0.2200	KS	BAD	CORRECT
0.2200	KS	BAD	INCORRECT
0.2200	KS	BAD	CORRECT
0.2200	KS	BAD	INCORRECT
0.5000	Default	BAD	CORRECT
0.5000	Default	BAD	INCORRECT
0.5000	Default	BAD	CORRECT
0.5000	Default	BAD	INCORRECT

Event	Value	Training Frequency	Validation Frequency
1	True Positive	748	
1	False Negative	441	
0	True Negative	3,881	
0	False Positive	890	
1	True Positive	277	
1	False Negative	912	
0	True Negative	4,666	
0	False Positive	105	

Test Frequency	Training Percentage	Validation Percentage	Test Percentage
	62.9100		
	37.0900		
	81.3456		
	18.6544		
	23.2969		
	76.7031		
	97.7992		

Test Frequency	Training Percentage	Validation Percentage	Test Percentage
	2.2008		

Properties

Property Name	Property Value
actFunc1	TANH
actFunc10	TANH
actFunc2	TANH
actFunc3	TANH
actFunc4	TANH
actFunc5	TANH
actFunc6	TANH
actFunc7	TANH
actFunc8	TANH
actFunc9	TANH
actFuncAll	TANH
annealingRate	0.0000
atAppendLookup	false
atCreateHistory	false
atHistoryLibUri	
atHistoryTblName	
atLeaveAutotuneOn	false
atLookupTableUri	
atMaxBayes	100
atMaxEval	50
atMaxIter	5
atMaxTime	60
atObjectiveInt	ASE
atObjectiveNom	KS
atPopSize	10
atSampleSize	50
atSearchMethod	GA
atTrainProp	0.7000

Property Name	Property Value
atUpdateProperties	false
atUseLookup	false
atValidFold	5
atValidMethod	PARTITION
atValidProp	0.3000
atannealingRate	true
atannealingRateInit	0.0010
atannealingRateLB	0.0000
atannealingRateUB	0.1000
atlearningRate	true
atlearningRateInit	0.0010
atlearningRateLB	0
atlearningRateUB	0.1000
atnhidden	true
atnhiddenInit	1
atnhiddenLB	0
atnhiddenUB	2
atnunitsInit	1
atnunitsLB	1
atnunitsUB	100
atweightDecay1	true
atweightDecay1Init	0
atweightDecay1LB	0
atweightDecay1UB	10
atweightDecay2	true
atweightDecay2Init	0
atweightDecay2LB	0
atweightDecay2UB	10
autotune_enabled	false

Property Name	Property Value
binaryProbCutoff	0.5000
codeLocation	mlearning
dataMiningVersion	V2024.03
directConn	false
dnnAlg	ADAM
dnnBeta1	0.9000
dnnBeta2	0.9990
dnnGamma	0.1000
dnnLRPolicy	FIXED
dnnMaxEpochs	10
dnnMomentum	0.9000
dnnPower	0.7500
dnnStepSize	10
earlyStop	true
exactPctlLift	true
explainFidelity	false
explainInfo	false
fullDatasetReconstitution	false
goal	0
hidden1	50
hidden10	50
hidden2	50
hidden3	50
hidden4	50
hidden5	50
hidden6	50
hidden7	50
hidden8	50

Property Name	Property Value
hidden9	50
hiddenAll	true
hiddenAllNum	50
hiddenDropout	0
icePlots	false
inputDropout	0
inputStd	MIDRANGE
learningRate	0.0010
maxIter	300
maxNumShapVars	20
maxTime	0
miniBatchSize	50
missAsLevl	false
momentum	0
nBins	50
nHidden	1
numCorrections	6
numTries	1
optTech	AUTOMATIC
pdNumImportantInputs	5
pdObsSamples	1,000
pdPlots	false
performKernelShap	false
performLime	false
performVI	false
randomSeed	12,345
reportingOnly	false
seedId	12,345

Property Name	Property Value
sgdSeed	12,345
specifyRows	RANDOM
stagnation	5
targetAct	IDENTITY
targetError	NORMAL
targetStd	MIDRANGE
templateRevision	5
train	true
truncateLI	5
truncateUI	95
useLocking	false
userProbCutoff	false
weightDecay	0.1000
weightDecay1	0

Output

The SAS System

The NNET Procedure

Model Information	
Model	Neural Net
Number of Observations Used	5960
Number of Observations Read	5960
Target/Response Variable	BAD
Number of Nodes	105
Number of Input Nodes	53
Number of Output Nodes	2
Number of Hidden Nodes	50
Number of Hidden Layers	1
Number of Weight Parameters	2700
Number of Bias Parameters	52
Architecture	MLP
Seed for Initial Weight	12345
Optimization Technique	LBFGS
Number of Neural Nets	1
Objective Value	1.8704884177

Iteration History

Iteration Number	Objective Function	Norm of Gradient	Loss	Step Size	Norm L1	Norm L2	Norm Maximum	Fit Error
1	2.828715	1.245243	2.771262	0	26.02551	0.574532	0.019722	0.357383
2	2.095523	0.333406	2.039755	0.800716	23.88770	0.557682	0.069953	0.199497
3	2.041216	0.102133	1.983893	1	21.40122	0.573227	0.104699	0.199497
4	2.031229	0.097229	1.979850	1	18.45064	0.513786	0.095949	0.199497
5	2.017152	0.142773	1.970823	1	11.27245	0.463293	0.094434	0.199497
6	2.012425	0.153358	1.959148	1	10.94163	0.532763	0.117470	0.199497
7	1.944301	0.202607	1.778348	1.827910	50.21701	1.659532	0.360204	0.197483
8	1.942071	0.202038	1.768727	0.007442	52.73521	1.733442	0.375402	0.196309
9	1.937519	0.155504	1.752965	0.376944	56.25377	1.845537	0.399326	0.192450
10	1.933824	0.138011	1.740618	1	58.61653	1.932063	0.419935	0.187919
11	1.918188	0.116017	1.707318	1	61.17429	2.108695	0.473493	0.181879
12	1.898463	0.111038	1.667850	1	60.74290	2.306135	0.538266	0.176510
13	1.880771	0.090897	1.641262	1	57.98002	2.395092	0.560063	0.173322
14	1.874838	0.116793	1.640368	1	54.01392	2.344703	0.537039	0.172148
15	1.873268	0.108612	1.635787	1	55.46663	2.374814	0.533502	0.172819
16	1.872666	0.115308	1.634150	1	55.62496	2.385164	0.534524	0.172148
17	1.872326	0.100461	1.632286	1	55.76887	2.400399	0.537066	0.172148
18	1.872196	0.099016	1.631809	1	55.93030	2.403869	0.536051	0.171980
19	1.871604	0.097872	1.631078	1	55.81981	2.405262	0.531644	0.172148
20	1.871226	0.099663	1.630458	1	55.61664	2.407683	0.529070	0.171980
21	1.870993	0.099852	1.630218	1	55.63225	2.407749	0.526346	0.171644
22	1.870829	0.100605	1.629645	1	55.70008	2.411842	0.526742	0.171477
23	1.870717	0.102387	1.629009	1	55.88631	2.417077	0.527332	0.171477
24	1.870662	0.098817	1.628423	1	56.10608	2.422387	0.526913	0.171141
25	1.870615	0.098882	1.628031	1	56.41354	2.425841	0.525046	0.170638
26	1.870573	0.097643	1.627708	1	56.90459	2.428647	0.517178	0.170973
27	1.870728	0.127060	1.628080	0.402145	56.83819	2.426482	0.516815	0.170805
28	1.870531	0.100428	1.627732	1	56.88479	2.427986	0.517006	0.170638
29	1.870516	0.099701	1.627710	5	56.88976	2.428060	0.516714	0.170805
30	1.870504	0.099490	1.627714	0.309229	56.89082	2.427895	0.516269	0.170973
31	1.870514	0.106933	1.627897	3.122834	56.86719	2.426166	0.514520	0.170302
32	1.870488	0.100957	1.627802	1	56.88045	2.426864	0.514959	0.170638
33	1.870488	0.100957	1.627802	1.12E-12	56.88045	2.426864	0.514959	0.170638

The optimization achieved the desired objective value.

Predicted Probability Variables	
LevName	Variable
1	P_BAD1
0	P_BAD0

Predicted Target Variable
Variable
I_BAD