

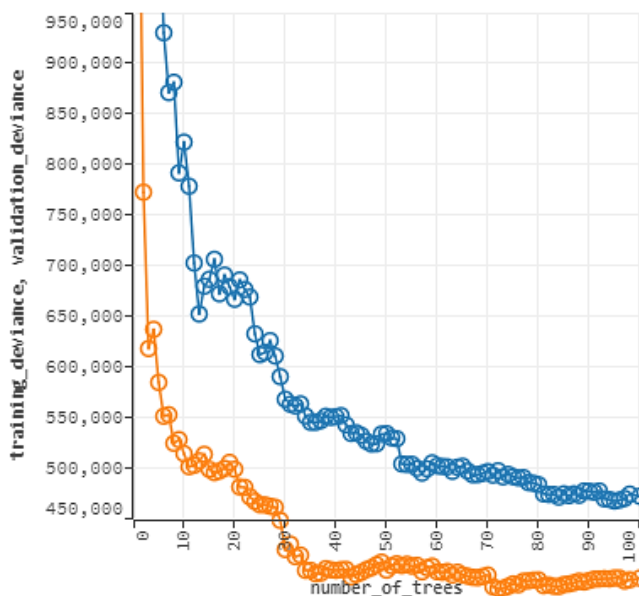
Compiled

Model	Distributed Random Forest	Deep Learning (MLP)	Generalized Linear Modeling
MSE	392222.702736	267682.987863	1639135.939137
RMSE	626.276858	517.380892	1280.287444
R2	0.853989	0.900375	0.389952
mean_residual_deviance	392222.702736	267682.987863	1639135.939137
MAE	248.689647	285.691315	578.138688
VARIABLE IMPORTANCES	Scaled Importance		
Rd_Pedes	1.0	0.8141	0.9080
Weekend	0.8463	1.0	1.0
Comm_lur	0.7939	0.8791	0.8106
Summer	0.6872	0.8084	0.7314
Income	0.5513	0.8076	0.6128
AreaProx	0.3964	0.7151	0.7244
Dist_center	0.3737	0.9129	0.6934
Other_lur	0.2135	0.8992	0.4709
Resi_lur	0.1484	0.8532	0.2346
Public_lur	0.1232	0.7583	0.1002
Winter	0.1228	0.8553	0.4125
T_Enable	0.1184	0.7759	0.0811
LUP_Other	0.1067	0.8002	0.1684
T_Hinder	0.0875	0.8133	0.2616
Recre_lur	0.0766	0.8438	0.0086
Rd_Sec	0.0739	0.7143	0.1640
Rd_Ter	0.0652	0.8330	0.0887
Rd_Cyc	0.0616	0.8205	0.2237
Edu_lur	0.0393	0.8435	0.1518
Spring	0.0374	0.8104	0.1637
Health_lur	0.0240	0.8680	0.0281
Rd_Pri	0.0166	0.8043	0.1571

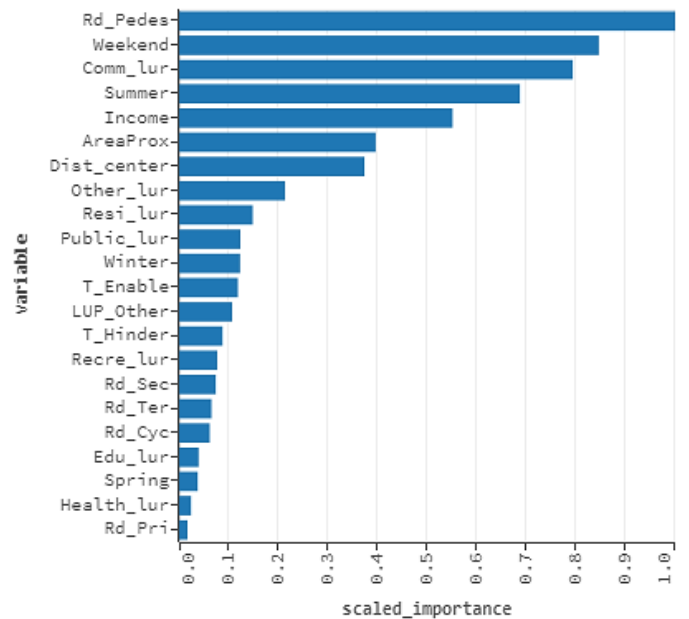
Distributed Random Forest

<i>training_frame</i>	Stodata.hex_0.80	Id of the training data frame.
<i>validation_frame</i>	Stodata.hex_0.20	Id of the validation data frame.
<i>fold_assignment</i>		Cross-validation fold assignment scheme, if fold_column is not specified. The 'Stratified' option will stratify the folds based on the response variable, for classification problems.
<i>response_column</i>	trip_count	Response variable column.
<i>ignored_columns</i>	FID_grid, geometry, x, y	Names of columns to ignore for training.
<i>ntrees</i>	100	Number of trees.
<i>r2_stopping</i>	1.7976931348623157e+308	r2_stopping is no longer supported and will be ignored if set - please use stopping_rounds, stopping_metric and stopping_tolerance instead. Previous version of H2O would stop making trees when the R^2 metric equals or exceeds this
<i>stopping_metric</i>		Metric to use for early stopping (AUTO: logloss for classification, deviance for regression and anomaly_score for Isolation Forest). Note that custom and custom_increasing can only be used in GBM and DRF with the Python client.
<i>seed</i>	3119900396890848489	Seed for pseudo random number generator (if applicable)
<i>histogram_type</i>	UniformAdaptive	What type of histogram to use for finding optimal split points
<i>categorical_encoding</i>	Enum	Encoding scheme for categorical features
<i>calibration_method</i>	PlattScaling	Calibration method to use
<i>distribution</i>	gaussian	Distribution function

▼ SCORING HISTORY - DEVIANCE



▼ VARIABLE IMPORTANCES



OUTPUT

	Validation Metric	Training Metric
<i>model</i>	drf-8fdf7da2-9231-4a07-b0c7-ce085b8b629a	drf-8fdf7da2-9231-4a07-b0c7-ce085b8b629a
<i>model_checksum</i>	-8261231867618916028	-8261231867618916028
<i>frame</i>	Stodata.hex 0.20	Stodata.hex 0.80
<i>frame_checksum</i>	-8729099454355841722	7843239400962262
<i>description</i>	•	Metrics reported on Out-Of-Bag training samples
<i>model_category</i>	Regression	Regression
<i>scoring_time</i>	1691595112601	1691595112601
<i>predictions</i>	•	•
<i>MSE</i>	392222.702736	472993.047778
<i>RMSE</i>	626.276858	687.744900
<i>nobs</i>	595	2395
<i>custom_metric_name</i>	•	•
<i>custom_metric_value</i>	0	0
<i>r2</i>	0.853989	0.823963
<i>mean_residual_deviance</i>	392222.702736	472993.047778
<i>mae</i>	248.689647	248.421668
<i>rmsle</i>	0.920942	0.948181

OUTPUT - VARIABLE IMPORTANCES

<i>variable</i>	<i>relative_importance</i>	<i>scaled_importance</i>	<i>percentage</i>
Rd Pedes	73046933504.0	1.0	0.1677
Weekend	61816143872.0	0.8463	0.1419
Comm_lur	57988395008.0	0.7939	0.1331
Summer	50200154112.0	0.6872	0.1152
Income	40272003072.0	0.5513	0.0924
AreaProx	28953833472.0	0.3964	0.0665
Dist_center	27299696640.0	0.3737	0.0627
Other_lur	15592902656.0	0.2135	0.0358
Resi_lur	10839425024.0	0.1484	0.0249
Public_lur	8997611520.0	0.1232	0.0207
Winter	8966690816.0	0.1228	0.0206
T_Enable	8651498496.0	0.1184	0.0199
LUP_Other	7793270784.0	0.1067	0.0179
T_Hinder	6389305856.0	0.0875	0.0147
Recre_lur	5595463168.0	0.0766	0.0128
Rd_Sec	5401158656.0	0.0739	0.0124
Rd_Ter	4760001024.0	0.0652	0.0109
Rd_Cyc	4501403136.0	0.0616	0.0103
Edu_lur	2868453888.0	0.0393	0.0066
Spring	2734832128.0	0.0374	0.0063
Health_lur	1754786944.0	0.0240	0.0040
Rd_Pri	1213055232.0	0.0166	0.0028

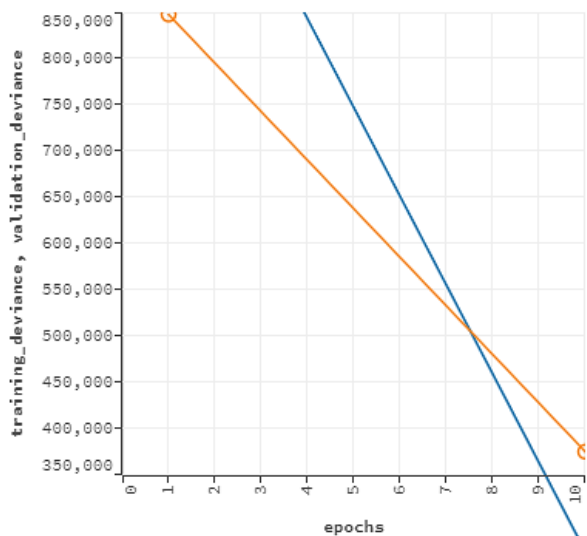
MODEL SUMMARY

<i>number_of_trees</i>	100
<i>number of internal trees</i>	100
<i>model_size_in_bytes</i>	1674979
<i>min_depth</i>	20
<i>max_depth</i>	20
<i>mean_depth</i>	20.0
<i>min_leaves</i>	1253
<i>max_leaves</i>	1386
<i>mean_leaves</i>	1325.8800

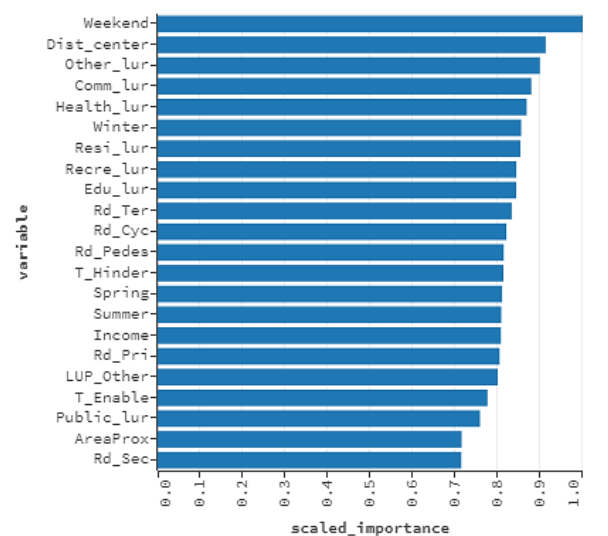
Deep Learning

Parameter	Value	Description
model_id	deeplearning-254fddab-83cf-4cc7-b647-8bedeafdcdd9	Destination id for this model; auto-generated if not specified.
training_frame	Stodata.hex_0.80	Id of the training data frame.
validation_frame	Stodata.hex_0.20	Id of the validation data frame.
fold_assignment		Cross-validation fold assignment scheme, if fold_column is not specified. The 'Stratified' option will stratify the folds based on the response variable, for classification problems.
response_column	trip_count	Response variable column.
ignored_columns	FID_grid, geometry, x, y	Names of columns to ignore for training.
hidden	200, 200	Hidden layer sizes (e.g. [100, 100]).
seed	-1269310814138902646	Seed for random numbers (affects sampling) - Note: only reproducible when running single threaded.
distribution	gaussian	Distribution function
stopping_metric	deviance	Metric to use for early stopping (AUTO: logloss for classification, deviance for regression and anomaly_score for Isolation Forest). Note that custom and custom_increasing can only be used in GBM and DRF with the Python client.
categorical_encoding	OneHotInternal	Encoding scheme for categorical features

▼ SCORING HISTORY - DEVIANCE



▼ VARIABLE IMPORTANCES



Output - status of neuron layers (predicting trip_count, regression, gaussian distribution, quadratic loss, 45 001 weights/biases, 541,1 kb, 23 950 training samples, mini-batch size 1)

layer	units	type	dropout	l1	l2	mean_rate	rate_rms	momentum	mean_weight	weight_rms	mean_bias	bias_rms
1	22	Input	0									
2	200	Rectifier	0	0	0	0.0186	0.0108	0	0.0007	0.0992	0.4895	0.0487
3	200	Rectifier	0	0	0	0.0839	0.1365	0	-0.0183	0.0789	0.9807	0.0136
4	1	Linear		0	0	0.0010	0.0009	0	0.0279	0.0696	0.0117	0.0

OUTPUT - TRAINING_METRICS

<i>model</i>	deeplearning-254fddab-83cf-4cc7-b647-8bedeafdcdd9	deeplearning-254fddab-83cf-4cc7-b647-8bedeafdcdd9
<i>model_checksum</i>	7302145103267320824	7302145103267320824
<i>frame</i>	Stodata.hex 0.80	Stodata.hex 0.20
<i>frame_checksum</i>	7843239400962262	-8729099454355841722
<i>description</i>	Metrics reported on full training frame	Metrics reported on full validation frame
<i>model_category</i>	Regression	Regression
<i>scoring_time</i>	1691595698718	1691595698722
<i>predictions</i>	.	.
<i>MSE</i>	267682.987863	375295.859174
<i>RMSE</i>	517.380892	612.613956
<i>nobs</i>	2395	595
<i>custom_metric_name</i>	.	.
<i>custom_metric_value</i>	0	0
<i>r2</i>	0.900375	0.860290
<i>mean_residual_deviance</i>	267682.987863	375295.859174
<i>mae</i>	285.691315	346.640761
<i>rmsle</i>	NaN	NaN

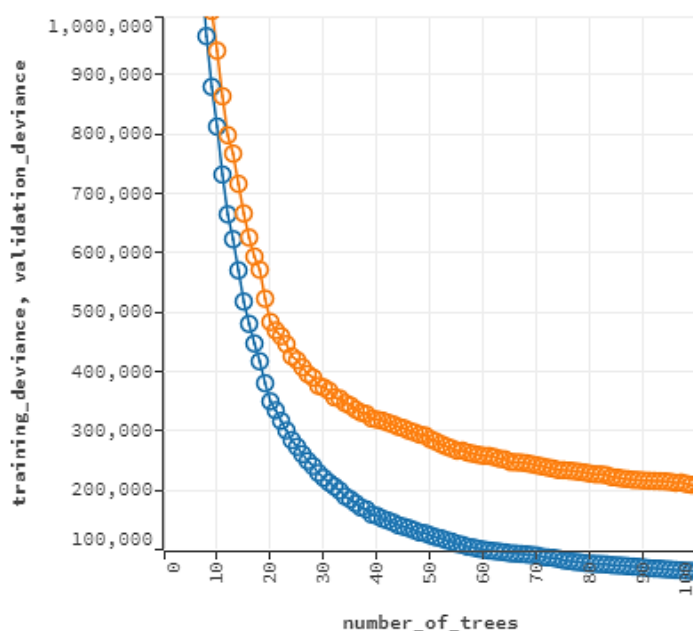
OUTPUT - VARIABLE IMPORTANCES

<i>variable</i>	<i>relative_importance</i>	<i>scaled_importance</i>	<i>percentage</i>
Weekend	1.0	1.0	0.0549
Dist_center	0.9129	0.9129	0.0501
Other_lur	0.8992	0.8992	0.0493
Comm_lur	0.8791	0.8791	0.0482
Health_lur	0.8680	0.8680	0.0476
Winter	0.8553	0.8553	0.0469
Resi_lur	0.8532	0.8532	0.0468
Recre_lur	0.8438	0.8438	0.0463
Edu_lur	0.8435	0.8435	0.0463
Rd_Ter	0.8330	0.8330	0.0457
Rd_Cyc	0.8205	0.8205	0.0450
Rd_Pedes	0.8141	0.8141	0.0447
T_Hinder	0.8133	0.8133	0.0446
Spring	0.8104	0.8104	0.0445
Summer	0.8084	0.8084	0.0443
Income	0.8076	0.8076	0.0443
Rd_Pri	0.8043	0.8043	0.0441
LUP_Other	0.8002	0.8002	0.0439
T_Enable	0.7759	0.7759	0.0426
Public_lur	0.7583	0.7583	0.0416
AreaProx	0.7151	0.7151	0.0392
Rd_Sec	0.7143	0.7143	0.0392

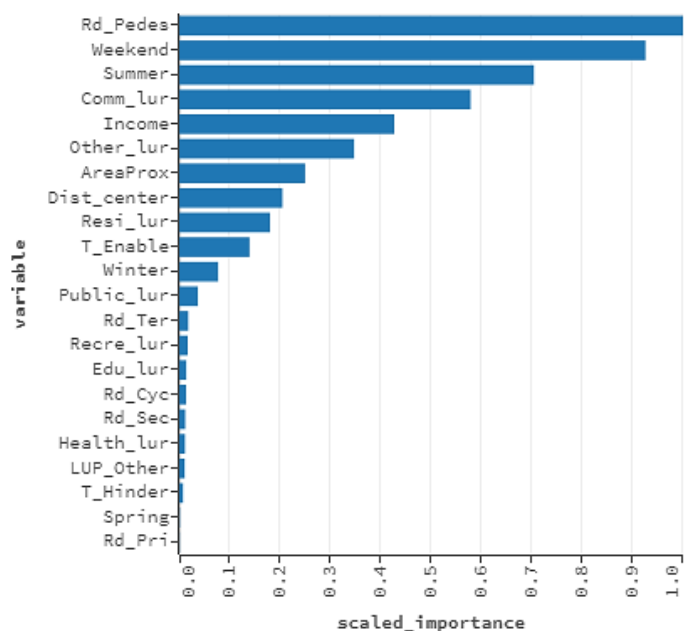
Gradient Boosting Machine

Parameter	Value	Description
model_id	gbm-f778ffc9-1b14-4458-82ed-7d2e89994cd8	Destination id for this model; auto-generated if not specified.
training_frame	Stodata.hex_0.80	Id of the training data frame.
validation_frame	Stodata.hex_0.20	Id of the validation data frame.
fold_assignment		Cross-validation fold assignment scheme, if fold_column is not specified. The 'Stratified' option will stratify the folds based on the response variable, for classification problems.
response_column	trip_count	Response variable column.
ignored_columns	FID_grid, geometry, x, y	Names of columns to ignore for training.
ntrees	100	Number of trees.
r2_stopping	1.7976931348623157e+308	r2_stopping is no longer supported and will be ignored if set - please use stopping_rounds, stopping_metric and stopping_tolerance instead. Previous version of H2O would stop making trees when the R^2 metric equals or exceeds this
stopping_metric		Metric to use for early stopping (AUTO: logloss for classification, deviance for regression and anomaly_score for Isolation Forest). Note that custom and custom_increasing can only be used in GBM and DRF with the Python client.
seed	2657379607235522297	Seed for pseudo random number generator (if applicable)
distribution	gaussian	Distribution function
histogram_type	UniformAdaptive	What type of histogram to use for finding optimal split points
max_abs_leafnode_pred	1.7976931348623157e+308	Maximum absolute value of a leaf node prediction
categorical_encoding	Enum	Encoding scheme for categorical features
calibration_method	PlattScaling	Calibration method to use

▼ SCORING HISTORY - DEVIANCE



▼ VARIABLE IMPORTANCES



TRAINING_METRICS		VALIDATION_METRICS
model	gbm-f778ffc9-1b14-4458-82ed-7d2e89994cd8	gbm-f778ffc9-1b14-4458-82ed-7d2e89994cd8
model_checksum	-3991065439983705140	-3991065439983705140
frame	Stodata.hex 0.80	Stodata.hex 0.20
frame_checksum	7843239400962262	-8729099454355841722
description	.	.
model_category	Regression	Regression
scoring_time	1691596546260	1691596546260
predictions	.	.
MSE	65156.866410	211027.862995
RMSE	255.258431	459.377691
nobs	2395	595
custom_metric_name	.	.
custom_metric_value	0	0
r2	0.975750	0.921441
mean_residual_deviance	65156.866410	211027.862995
mae	127.346813	200.760325
rmsle	NaN	NaN

OUTPUT - VARIABLE IMPORTANCES

variable	relative_importance	scaled_importance	percentage
Rd Pedes	6642171904.0	1.0	0.2010
Weekend	6148124160.0	0.9256	0.1860
Summer	4677260288.0	0.7042	0.1415
Comm_lur	3841497344.0	0.5783	0.1162
Income	2835324416.0	0.4269	0.0858
Other_lur	2305684992.0	0.3471	0.0698
AreaProx	1658233472.0	0.2497	0.0502
Dist_center	1360136704.0	0.2048	0.0412
Resi_lur	1196895744.0	0.1802	0.0362
T_Enable	926326912.0	0.1395	0.0280
Winter	512441632.0	0.0771	0.0155
Public_lur	246100784.0	0.0371	0.0074
Rd_Ter	117782928.0	0.0177	0.0036
Recre_lur	110605080.0	0.0167	0.0033
Edu_lur	91861568.0	0.0138	0.0028
Rd_Cyc	89945184.0	0.0135	0.0027
Rd_Sec	78335016.0	0.0118	0.0024

Health_lur	74468752.0	0.0112	0.0023
LUP_Other	71212376.0	0.0107	0.0022
T_Hinder	49654232.0	0.0075	0.0015
Spring	8336166.5000	0.0013	0.0003
Rd_Pri	5337236.5000	0.0008	0.0002

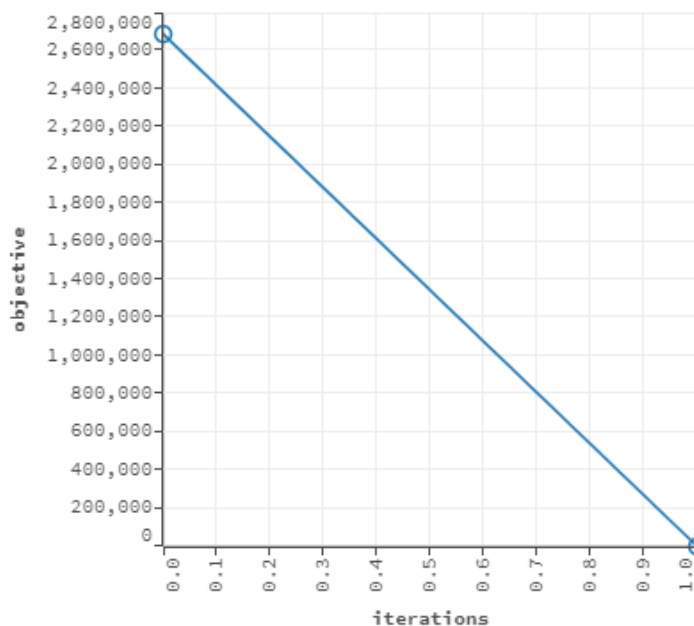
OUTPUT - MODEL SUMMARY

number_of_trees	100
number_of_internal_trees	100
model_size_in_bytes	33482
min_depth	5
max_depth	5
mean_depth	5.0
min_leaves	11
max_leaves	30
mean_leaves	21.9300

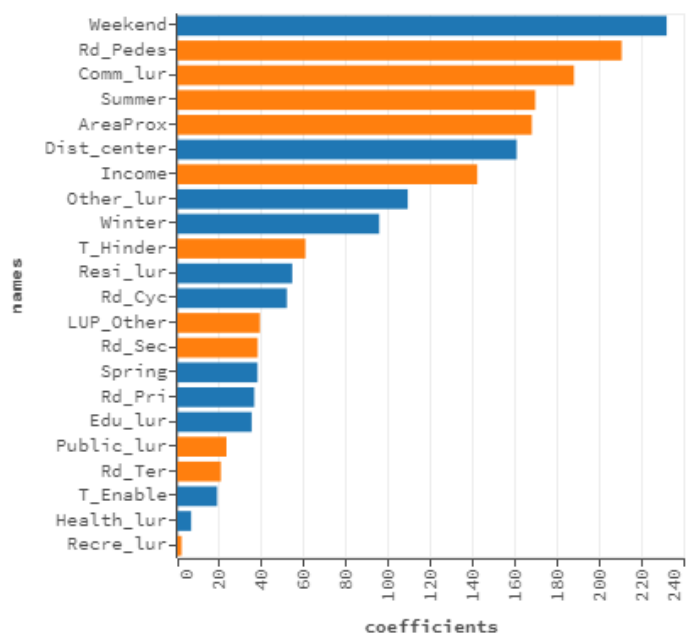
Generalized Linear Modeling

Parameter	Value	Description
model_id	glm-e9d509ab-d803-42d9-8133-c52c6dc1416d	Destination id for this model; auto-generated if not specified.
training_frame	Stodata.hex_0.80	Id of the training data frame.
validation_frame	Stodata.hex_0.20	Id of the validation data frame.
seed	4734591805501708305	Seed for pseudo random number generator (if applicable)
fold_assignment		Cross-validation fold assignment scheme, if fold_column is not specified. The 'Stratified' option will stratify the folds based on the response variable, for classification problems.
response_column	trip_count	Response variable column.
ignored_columns	FID_grid, geometry, x, y	Names of columns to ignore for training.
family	gaussian	Family. Use binomial for classification with logistic regression, others are for regression problems.
solver	IRLSM	AUTO will set the solver based on given data and the other parameters. IRLSM is fast on problems with small number of predictors and for lambda-search with L1 penalty, L_BFGS scales better for datasets with many columns.
alpha	0.5	Distribution of regularization between the L1 (Lasso) and L2 (Ridge) penalties. A value of 1 for alpha represents Lasso regression, a value of 0 produces Ridge regression, and anything in between specifies the amount of mixing between the two. Default value of alpha is 0 when SOLVER = 'L-BFGS'; 0.5 otherwise.
lambda	1.448612553736424	Regularization strength
max_iterations	50	Maximum number of iterations
objective_epsilon	0.0001	Converge if objective value changes less than this. Default (of -1.0) indicates: If lambda_search is set to True the value of objective_epsilon is set to .0001. If the lambda_search is set to False and lambda is equal to zero, the value of objective_epsilon is set to .000001, for any other value of lambda the default value of objective_epsilon is set to .0001.
gradient_epsilon	0.0001	Converge if objective changes less (using L-infinity norm) than this, ONLY applies to L-BFGS solver. Default (of -1.0) indicates: If lambda_search is set to False and lambda is equal to zero, the default value of gradient_epsilon is equal to .000001, otherwise the default value is .0001. If lambda_search is set to True, the conditional values above are 1E-8 and 1E-6 respectively.
link	identity	Link function.
lambda_min_ratio	0.0001	Minimum lambda used in lambda search, specified as a ratio of lambda_max (the smallest lambda that drives all coefficients to zero). Default indicates: if the number of observations is greater than the number of variables, then lambda_min_ratio is set to 0.0001; if the number of observations is less than the number of variables, then lambda_min_ratio is set to 0.01.
max_active_predictors	5000	Maximum number of active predictors during computation. Use as a stopping criterion to prevent expensive model building with many predictors. Default indicates: If the IRLSM solver is used, the value of max_active_predictors is set to 5000 otherwise it is set to 100000000.
obj_reg	0.0004175365344467641	Likelihood divider in objective value computation, default (of -1.0) will set it to 1/nobs

SCORING HISTORY



STANDARDIZED COEFFICIENT MAGNITUDES



	TRAINING_METRICS	VALIDATION_METRICS
<i>model</i>	glm-e9d509ab-d803-42d9-8133-c52c6dc1416d	glm-e9d509ab-d803-42d9-8133-c52c6dc1416d
<i>model_checksum</i>	2991274418473774048	2991274418473774048
<i>frame</i>	Stodata.hex 0.80	Stodata.hex 0.20
<i>frame_checksum</i>	7843239400962262	-8729099454355841722
<i>description</i>	.	.
<i>model_category</i>	Regression	Regression
<i>scoring_time</i>	1691596941244	1691596941259
<i>predictions</i>	.	.
<i>MSE</i>	1639135.939137	1416942.536718
<i>RMSE</i>	1280.287444	1190.353954
<i>nobs</i>	2395	595
<i>custom_metric_name</i>	.	.
<i>custom_metric_value</i>	0	0
<i>r2</i>	0.389952	0.472520
<i>mean_residual_deviance</i>	1639135.939137	1416942.536718
<i>mae</i>	578.138688	603.798470
<i>rmsle</i>	NaN	NaN
<i>residual_deviance</i>	3925730574.233094	843080809.347389
<i>null_deviance</i>	6435120356.381628	1599977966.494752
<i>AIC</i>	41116.398682	10164.123974
<i>loglikelihood</i>	0	0
<i>null_degrees_of_freedom</i>	2394	594
<i>residual_degrees_of_freedom</i>	2372	572

OUTPUT - VARIABLE IMPORTANCES

<i>variable</i>	<i>relative_importance</i>	<i>scaled_importance</i>	<i>percentage</i>
Weekend	231.2676	1.0	0.1220
Rd_Pedes	210.0009	0.9080	0.1108
Comm_lur	187.4547	0.8106	0.0989
Summer	169.1511	0.7314	0.0892
AreaProx	167.5223	0.7244	0.0884
Dist_center	160.3680	0.6934	0.0846
Income	141.7308	0.6128	0.0748
Other_lur	108.9023	0.4709	0.0575
Winter	95.3936	0.4125	0.0503
T_Hinder	60.4890	0.2616	0.0319
Resi_lur	54.2612	0.2346	0.0286
Rd_Cyc	51.7346	0.2237	0.0273
LUP_Other	38.9500	0.1684	0.0205
Rd_Sec	37.9236	0.1640	0.0200
Spring	37.8516	0.1637	0.0200
Rd_Pri	36.3311	0.1571	0.0192
Edu_lur	35.1116	0.1518	0.0185
Public_lur	23.1803	0.1002	0.0122
Rd_Ter	20.5205	0.0887	0.0108
T_Enable	18.7567	0.0811	0.0099
Health_lur	6.4965	0.0281	0.0034
Recre_lur	1.9892	0.0086	0.0010

OUTPUT - GLM MODEL (SUMMARY)

<i>family</i>	gaussian
<i>link</i>	identity
<i>regularization</i>	Elastic Net (alpha = 0.5, lambda = 1.4486)
<i>number_of_predictors_total</i>	22
<i>number_of_active_predictors</i>	22
<i>number_of_iterations</i>	1
<i>training_frame</i>	Stodata.hex_0.80

OUTPUT - COEFFICIENTS (GLM COEFFICIENTS)

<i>names</i>	<i>coefficients</i>	<i>standardized_coefficients</i>	<i>names</i>	<i>coefficients</i>	<i>sign</i>
Intercept	421.2983	765.0484	Weekend	231.2676	NEG
Comm_lur	988.2622	187.4547	Rd_Pedes	210.0009	POS
Income	0.3217	141.7308	Comm_lur	187.4547	POS
T_Enable	-51.7062	-18.7567	Summer	169.1511	POS
Rd_Pedes	0.1238	210.0009	AreaProx	167.5223	POS
Rd_Ter	0.0206	20.5205	Dist_center	160.3680	NEG
T_Hinder	181.3689	60.4890	Income	141.7308	POS
Resi_lur	-273.9361	-54.2612	Other_lur	108.9023	NEG
Health_lur	-105.6123	-6.4965	Winter	95.3936	NEG
AreaProx	0.0003	167.5223	T_Hinder	60.4890	POS
Other_lur	-543.4397	-108.9023	Resi_lur	54.2612	NEG
Edu_lur	-310.7642	-35.1116	Rd_Cyc	51.7346	NEG
LUP_Other	1.7237	38.9500	LUP_Other	38.9500	POS
Recre_lur	16.7866	1.9892	Rd_Sec	37.9236	POS
Rd_Cyc	-0.0880	-51.7347	Spring	37.8516	NEG
Public_lur	273.0107	23.1803	Rd_Pri	36.3311	NEG
Rd_Sec	0.0615	37.9236	Edu_lur	35.1116	NEG
Summer	378.3244	169.1511	Public_lur	23.1803	POS
Winter	-220.7189	-95.3936	Rd_Ter	20.5205	POS
Weekend	-462.5986	-231.2676	T_Enable	18.7567	NEG
Rd_Pri	-0.0941	-36.3311	Health_lur	6.4965	NEG
Spring	-90.6781	-37.8516	Recre_lur	1.9892	POS
Dist_center	-0.0921	-160.3680			