root

Go Up

Table of Contents

		10		•		r
н	Sinomia	11	'ont	1101	on oc	
ப	шошпа	\mathbf{L}	ω	ւսու	OII.EC	L

Binomial confusion matrix

BinomialLogisticRegression.ecl

Binomial logistic regression using iteratively re-weighted least squares

Confusion.ecl

Detail confusion records to compare actual versus predicted response variable values

Constants.ecl

DataStats.ecl

Information about the datasets

Deviance_Analysis.ecl

Compare deviance information for an analysis of deviance

Deviance Detail.ecl

Detail deviance for each observation

dimm.ecl

Matrix multiply when either A or B is a diagonal and is passed as a vector

Distributions.ecl

ExtractBeta.ecl

Extract the beta values form the model dataset

 $ExtractBeta_CI.ecl$

Extract the beta values form the model dataset

ExtractBeta_pval.ecl

Extract the beta values form the model dataset

ExtractReport.ecl

Extract Report records from model
LogitPredict.ecl
Predict the category values with the logit function and the supplied beta coefficients
LogitScore.ecl
Calculate the score using the logit function and the supplied beta coefficients
Model_Deviance.ecl
Model Deviance
Null_Deviance.ecl
Deviance for the null model, that is, a model with only an intercept
Types.ecl
IRLS
performance
Tests
validation

Binomial Confusion

Go Up

IMPORTS

 $ML_Core.Types \mid Types \mid$

DESCRIPTIONS

FUNCTION BinomialConfusion

DATASET(Types.Binomial_Confusion_Summary)	BinomialConfusion
(DATASET(Core_Types.Confusion_Detail) d)	

Binomial confusion matrix. Work items with multinomial responses are ignored by this function. The higher value lexically is considered to be the positive indication.

PARAMETER $\underline{\mathbf{d}}$ confusion detail for the work item and classifier

RETURN confusion matrix for a binomial classifier

BinomialLogisticRegression

Go Up

IMPORTS

Constants | ML_Core.Interfaces | ML_Core.Types |

DESCRIPTIONS

MODULE BinomialLogisticRegression

BinomialLogisticRegression

(UNSIGNED max_iter=200, REAL8 epsilon=Constants.default_epsilon, REAL8 ridge=Constants.default_ridge)

Binomial logistic regression using iteratively re-weighted least squares.

PARAMETER max_iter maximum number of iterations to try

PARAMETER epsilon the minimum change in the Beta value estimate to continue

PARAMETER <u>ridge</u> a value to populate a diagonal matrix that is added to a matrix help assure that the matrix is invertible.

Children

- 1. GetModel: Calculate the model to fit the observation data to the observed classes
- 2. Classify: Classify the observations using a model
- 3. Report: Report the confusion matrix for the classifier and training data

FUNCTION GetModel

BinomialLogisticRegression \

DATASET(Types.Layout_Model)	$\operatorname{GetModel}$
(DATASET(Types.NumericField) DATASET(Types.DiscreteField)	•

Calculate the model to fit the observation data to the observed classes.

PARAMETER observations the observed explanatory values

PARAMETER <u>classifications</u> the observed classification used to build the model

RETURN the encoded model

OVERRIDE True

FUNCTION Classify

 $BinomialLogisticRegression \setminus$

DATASET(Types.Classify_Result)	Classify
(DATASET(Types.Layout_Model) mo DATASET(Types.NumericField) new	

Classify the observations using a model.

PARAMETER model The model, which must be produced by a corresponding getModel function.

PARAMETER new_observations observations to be classified

RETURN Classification with a confidence value

OVERRIDE True

FUNCTION Report

BinomialLogisticRegression \

```
DATASET(Types.Confusion_Detail) Report

(DATASET(Types.Layout_Model) model,
DATASET(Types.NumericField) observations,
DATASET(Types.DiscreteField) classifications)
```

Report the confusion matrix for the classifier and training data.

PARAMETER model the encoded model

PARAMETER observations the explanatory values.

PARAMETER <u>classifications</u> the classifications associated with the observations

RETURN the confusion matrix showing correct and incorrect results

OVERRIDE True

Confusion

Go Up

IMPORTS

ML_Core | ML_Core.Types | Types |

DESCRIPTIONS

FUNCTION Confusion

DATASET(Confusion_Detail)	Confusion
(DATASET(DiscreteField) de	pendents, DATASET(DiscreteField)
predicts)	

Detail confusion records to compare actual versus predicted response variable values.

PARAMETER dependents the original response values

PARAMETER predicts the predicted responses

RETURN confusion counts by predicted and actual response values.

Constants

Go Up

DESCRIPTIONS

MODULE Constants

Constants

Children

- 1. limit_card
- 2. default_epsilon
- 3. default_ridge
- 4. local_cap
- 5. id_base
- 6. id_iters
- 7. id_delta
- 8. id_correct
- 9. id_incorrect
- $10. \ id_stat_set$
- 11. id_betas
- 12. id_betas_coef
- 13. id_betas_SE
- 14. base_builder
- 15. base_max_iter
- 16. base_epsilon

17. base_ind_	vars
18. base_dep_	_vars
19. base_obs	
20. builder_irl	ls_local
21. builder_irl	ls_global
22. builder_so	ftmax
ATTRIBUTI	limit_card
Constants \	
UNSIGNED2 lin	mit_card
ATTRIBUTI	default_epsilon
Constants \	
REAL8 defaul	lt_epsilon
uelaa.	орыхох
ATTRIBUTI	default_ridge
Carrata set s	
Constants \	

 $default_ridge$

REAL8

ATTRIBUTE local_cap Constants \ UNSIGNED4 local_cap ATTRIBUTE id_base Constants \ id_base ATTRIBUTE id_iters Constants \setminus id_iters

ATTRIBUTE id_delta

Constants \

id_delta

ATTRIBUTE id_correct

Constants \

id_correct
ATTRIBUTE id_incorrect
Constants \
id_incorrect
ATTRIBUTE id_stat_set
$\operatorname{Constants}\setminus$
id_stat_set
ATTRIBUTE id_betas
Constants \
id_betas
ATTRIBUTE id_betas_coef
Constants \
id_betas_coef

ATTRIBUTE id_betas_SE Constants \ id_betas_SE **ATTRIBUTE** base_builder Constants \ base_builder **ATTRIBUTE** base_max_iter Constants \ base_max_iter **ATTRIBUTE** base_epsilon Constants \ $base_epsilon$ **ATTRIBUTE** base_ind_vars

Constants \

base_ind_vars	
ATTRIBUTE has a day was	
ATTRIBUTE base_dep_vars	
$\operatorname{Constants} \setminus$	
Constants	
base_dep_vars	
ATTRIBUTE base_obs	
$\operatorname{Constants} \setminus$	
Constants	
base_obs	
ATTRIBUTE builder_irls_local	
Constants	
Constants \	
builder_irls_local	
ATTRIBUTE builder_irls_global	
Clonestanta	
Constants \	
builder_irls_global	\neg

ATTRIBUTE builder_softmax

						١
	α r	10	ta:	nt	C	١
$\mathbf{\circ}$	$\mathbf{o}_{\mathbf{I}}$	\mathbf{r}	υcu.	ш	JID.	

builder_softmax

DataStats

Go Up

IMPORTS

LogisticRegression.Types | LogisticRegression.Constants | ML_Core.Types |

DESCRIPTIONS

FUNCTION DataStats

DATASET(Types.Data_Info) DataStats

(DATASET(Core_Types.NumericField) indep,
DATASET(Core_Types.DiscreteField) dep, BOOLEAN
field_details=FALSE)

Information about the datasets. Without details the range for the x and y (independent and dependent) columns. Note that a column of all zero values cannot be distinguished from a missing column. When details are requested, the cardinality, minimum, and maximum values are returned. A zero cardinality is returned when the field cardinality exceeds the Constants.limit_card value.

PARAMETER indep data set of independent variables

PARAMETER dep data set of dependent variables

PARAMETER field_details Boolean directive to provide field level info

Deviance_Analysis

Go Up

IMPORTS

Types |

DESCRIPTIONS

FUNCTION Deviance_Analysis

DATASET(Types.AOD_Record) Deviance_Analysis

(DATASET(Types.Deviance_Record) proposed,
DATASET(Types.Deviance_Record) base)

Compare deviance information for an analysis of deviance.

PARAMETER proposed the proposed model

PARAMETER base the base model for comparison

RETURN the comparison of the deviance between the models

Deviance_Detail

Go Up

IMPORTS

ML_Core | ML_Core.Types | Types |

DESCRIPTIONS

FUNCTION Deviance_Detail

DATASET(Types.Observation_Deviance) Deviance_Detail

(DATASET(Core_Types.DiscreteField) dependents,
DATASET(Types.Raw Prediction) predicts)

Detail deviance for each observation.

PARAMETER dependents original dependent records for the model

PARAMETER predicts the predicted values of the response variable

RETURN the deviance information by observation and the log likelihood of the predicted result.

dimm

Go Up

IMPORTS

std.blas | std.BLAS.Types |

DESCRIPTIONS

EMBED dimm

```
Types.matrix_t dimm

(BOOLEAN transposeA, BOOLEAN transposeB, BOOLEAN diagonalA, BOOLEAN diagonalB, Types.dimension_t m, Types.dimension_t n,

Types.dimension_t k, Types.value_t alpha, Types.matrix_t A,

Types.matrix_t B, Types.value_t beta=0.0, Types.matrix_t C=[])
```

Matrix multiply when either A or B is a diagonal and is passed as a vector. alpha*op(A) op(B) + beta*C where op() is transpose

PARAMETER transpose true when transpose of A is used

PARAMETER transpose true when transpose of B is used

PARAMETER diagonal true when A is the diagonal matrix

PARAMETER diagonal true when B is the diagonal matrix

PARAMETER m number of rows in product

PARAMETER n number of columns in product

PARAMETER k number of columns/rows for the multiplier/multiplicand

PARAMETER alpha scalar used on A

PARAMETER <u>A</u> matrix A

PARAMETER B matrix B

PARAMETER beta scalar for matrix C

Distributions

Go Up

IMPORTS

ML_Core.Constants | ML_Core.Math |

DESCRIPTIONS

MODULE Distributions

Distributions

Children

- 1. Normal_CDF : Cumulative Distribution of the standard normal distribution, the probability that a normal random variable will be smaller than x standard deviations above or below the mean
- 2. Normal PPF: Normal Distribution Percentage Point Function
- 3. T_CDF: Students t distribution integral evaluated between negative infinity and x
- 4. T_PPF: Percentage point function for the T distribution
- 5. Chi2 CDF: The cumulative distribution function for the Chi Square distribution
- 6. Chi2_PPF: The Chi Squared PPF function

FUNCTION Normal_CDF

Distributions \

REAL8 Normal_CDF

(REAL8 x)

Cumulative Distribution of the standard normal distribution, the probability that a normal random variable will be smaller than x standard deviations above or below the mean. Taken from C/C++ Mathematical Algorithms for Scientists and Engineers, n. Shammas, McGraw-Hill, 1995

PARAMETER $\underline{\mathbf{x}}$ the number of standard deviations

FUNCTION Normal_PPF

Distributions \

REAL8 Normal_PPF

(REAL8 x)

Normal Distribution Percentage Point Function. Translated from C/C++ Mathematical Algorithms for Scientists and Engineers, N. Shammas, McGraw-Hill, 1995

PARAMETER $\underline{\mathbf{x}}$ probability

FUNCTION T_CDF

Distributions \

REAL8 T_CDF

(REAL8 x, REAL8 df)

Students t distribution integral evaluated between negative infinity and x. Translated from NIST SEL DATAPAC Fortran TCDF.f source

PARAMETER $\underline{\mathbf{x}}$ value of the evaluation

PARAMETER df degrees of freedom

FUNCTION T_PPF

Distributions \

```
REAL8 T_PPF

(REAL8 x, REAL8 df)
```

Percentage point function for the T distribution. Translated from NIST SEL DATAPAC Fortran TPPF.f source

FUNCTION Chi2_CDF

Distributions \

```
REAL8 Chi2_CDF

(REAL8 x, REAL8 df)
```

The cumulative distribution function for the Chi Square distribution. the CDF for the specified degrees of freedom. Translated from the NIST SEL DATAPAC Fortran subroutine CHSCDF.

FUNCTION Chi2_PPF

Distributions \

```
REAL8 Chi2_PPF

(REAL8 x, REAL8 df)
```

The Chi Squared PPF function. Translated from the NIST SEL DATAPAC Fortran subroutine CHSPPF.

ExtractBeta

Go Up

IMPORTS

ML_Core.Types | Types |

DESCRIPTIONS

FUNCTION ExtractBeta

ExtractBeta

(DATASET(Core_Types.Layout_Model) mod_ds)

Extract the beta values form the model dataset.

PARAMETER mod_ds the model dataset

RETURN a beta values as Model Coefficient records, zero as the constant term.

ExtractBeta_CI

Go Up

IMPORTS

ML_Core.Types | Types |

DESCRIPTIONS

FUNCTION ExtractBeta_CI

Extract the beta values form the model dataset.

PARAMETER mod_ds the model dataset

PARAMETER <u>level</u> the significance value for the intervals

RETURN the beta values with confidence intervals term.

ExtractBeta_pval

Go Up

IMPORTS

ML_Core.Types | Types |

DESCRIPTIONS

FUNCTION ExtractBeta_pval

DATASET(Types.pval_Model_Coef)	ExtractBeta_pval
(DATASET(Core_Types.Layout_Model) mod_ds)	

Extract the beta values form the model dataset.

PARAMETER mod_ds the model dataset

RETURN the beta values with p-values as Model Coefficient records, zero as the constant term.

ExtractReport

Go Up

IMPORTS

 $\operatorname{ML_Core.Types} \mid \operatorname{Types} \mid \operatorname{Constants} \mid$

DESCRIPTIONS

FUNCTION ExtractReport

DATASET(Types.Model_Report)	ExtractReport
(DATASET(Core_Types.Layout_Model) mod_ds)	

Extract Report records from model

PARAMETER mod_ds the model dataset

RETURN the model report dataset

LogitPredict

Go Up

IMPORTS

ML_Core.Types | Types |

DESCRIPTIONS

FUNCTION LogitPredict

DATASET(Classify_Result) LogitPredict

(DATASET(Model_Coef) coef, DATASET(NumericField)
independents)

Predict the category values with the logit function and the supplied beta coefficients.

PARAMETER coef the model beta coefficients

PARAMETER independents the observations

RETURN the predicted category values and a confidence score

LogitScore

Go Up

IMPORTS

 $\operatorname{ML_Core.Types} \mid \operatorname{Types} \mid$

DESCRIPTIONS

FUNCTION LogitScore

DATASET(Raw_Prediction) LogitScore

(DATASET(Model_Coef) coef, DATASET(NumericField)
independents)

Calculate the score using the logit function and the supplied beta coefficients.

PARAMETER coef the model beta coefficients

PARAMETER independents the observations

RETURN the raw prediction value

Model_Deviance

Go Up

IMPORTS

Types |

DESCRIPTIONS

FUNCTION Model_Deviance

DATASET(Types.Deviance_Record)	Model_Deviance
(DATASET(Types.Observation_Devi DATASET(Types.Model_Coef) mod)	

Model Deviance.

PARAMETER od observation deviance record

PARAMETER mod model co-efficients

RETURN model deviance

Null_Deviance

Go Up

IMPORTS

Types |

DESCRIPTIONS

FUNCTION Null_Deviance

DATASET(Types.Deviance_Record)	Null_Deviance
(DATASET(Types.Observation_Deviance) od)	

Deviance for the null model, that is, a model with only an intercept.

PARAMETER od Observation Deviance record set.

RETURN a data set of the null model deviances for each work item and classifier.

Types

Go Up

IMPORTS

ML_Core.Types |

DESCRIPTIONS

MODULE Types

Types

Children

- 1. t_Universe
- 2. Field_Desc
- 3. Data_Info
- 4. NumericField_U
- 5. DiscreteField_U
- 6. Layout_Column_Map
- 7. Classifier_Stats
- 8. Model_Report
- 9. Binomial_Confusion_Summary
- 10. Model_Coef
- 11. Confidence_Model_Coef

13	. Raw_Prediction
14	. Observation_Deviance
15	. Deviance_Record
16	. AOD_Record
A T	TDIDLITE
AI	TRIBUTE t_Universe
Тур	pes \
4	t_Universe
,	t_Oniverse
RE	CORD Field_Desc
Тур	pes \
]	Field_Desc
RE	CORD Data_Info
Tvr	pes \
+J1	
]	Data_Info
DE	CORD NumericField_U
ΚE	COND Numericrieia O
	

12. pval_Model_Coef

Types \setminus

NumericField_U
RECORD DiscreteField_U
ALCOND Discreter leid_C
$\mathrm{Types} \setminus$
DiscreteField_U
RECORD Layout_Column_Map
RECORD Layout_Column_Wap
$\mathrm{Types} \setminus$
-JP86 (
Layout_Column_Map
PECORD Classifier State
RECORD Classifier_Stats
$\mathrm{Types} \setminus$
Types (
Classifier_Stats
RECORD Model_Report
Types
Types \
Model_Report

RECORD Binomial_Confusion_Summary
Types \
Binomial_Confusion_Summary
RECORD Model_Coef
Types \
Model_Coef
RECORD Confidence_Model_Coef
Types \
Confidence_Model_Coef
RECORD pval_Model_Coef
Types \
pval_Model_Coef
RECORD Raw_Prediction
Types \

Raw_Prediction	
RECORD Observation_Deviance	
Types \	
Observation_Deviance	
RECORD Deviance_Record	
Types \	
Deviance_Record	
RECORD AOD_Record	
Types \	
AOD_Record	

IRLS

Go Up

Table of Contents

GetModel.ecl

Generate logistic regression model from training data

GetModel_global.ecl

Internal function to determine values for the model coefficients and selected statistics from building the model

$GetModel_local.ecl$

Internal function to determine values for the model co-efficients and selected stats from building the model

$\mathbf{GetModel}$

Go Up

IMPORTS

ML_Core | ML_Core.Types | Constants | Types | IRLS |

DESCRIPTIONS

FUNCTION GetModel

DATASET(Layout_Model) GetModel

(DATASET(NumericField) independents, DATASET(DiscreteField) dependents, UNSIGNED max_iter=200, REAL8 epsilon=Constants.default_epsilon, REAL8 ridge=Constants.default_ridge)

Generate logistic regression model from training data. The size of the inputs is used to determin which work items are processed with purely local operations (the data is moved once as necessary) or with global operations supporting a work item to use multiple nodes.

PARAMETER independents the independent values

PARAMETER dependents the dependent values.

PARAMETER <u>max_iter</u> maximum number of iterations to try

PARAMETER epsilon the minimum change in the Beta value estimate to continue

PARAMETER <u>ridge</u> a value to pupulate a diagonal matrix that is added to a matrix help assure that the matrix is invertible.

${\bf GetModel_global}$

Go Up

IMPORTS

ML Core | ML Core. Types | PBblas | PBblas. Types | Constants | Types |

DESCRIPTIONS

FUNCTION GetModel_global

DATASET(Layout_Model) GetModel_global

(DATASET(NumericField) independents, DATASET(DiscreteField)
dependents, UNSIGNED max_iter=200, REAL8
epsilon=Constants.default_epsilon, REAL8
ridge=Constants.default_ridge)

Internal function to determine values for the model coefficients and selected statistics from building the model.

PARAMETER independents the independent values

PARAMETER dependents the dependent values

PARAMETER max_iter maximum number of iterations to try

PARAMETER epsilon the minimum change in the Beta value estimate to continue

PARAMETER <u>ridge</u> a value to pupulate a diagonal matrix that is added to a matrix help assure that the matrix is invertible.

RETURN coefficient matrix plus model building statistics

1	1	٦
4	ι	,
_	•	,

$\begin{array}{c} {}_{\rm IRLS/} \\ {\bf GetModel \ local} \end{array}$

Go Up

IMPORTS

ML_Core | ML_Core.Types | Constants | Types | IRLS | std | std.blas |

DESCRIPTIONS

FUNCTION GetModel_local

DATASET(Layout_Model) | GetModel_local

(DATASET(NumericField) independents, DATASET(DiscreteField) dependents, UNSIGNED2 max_iter=200, REAL8 epsilon=Constants.default_epsilon, REAL8 ridge=Constants.default_ridge)

Internal function to determine values for the model co-efficients and selected stats from building the model.

PARAMETER independents the independent values

PARAMETER dependents the dependent values.

PARAMETER max_iter maximum number of iterations to try

PARAMETER epsilon the minimum change in the Beta value estimate to continue

PARAMETER <u>ridge</u> a value to populate a diagonal matrix that is added to a matrix help assure that the matrix is invertible.

RETURN coefficient matrix plus model building stats

1	റ
4	: 4

performance

Go Up

Table of Contents

RunBinomial.ecl

$\frac{\mathrm{performance}/}{\mathrm{RunBinomial}}$

Go Up

IMPORTS

 $\operatorname{ML_Core.Types}$ |

DESCRIPTIONS

ATTRIBUTE RunBinomial

RunBinomial

Tests

Go Up

Table of Contents

 ${\bf Check_Dist.ecl}$

$\begin{array}{c} {}^{\rm Tests/} \\ {\bf Check_Dist} \end{array}$

Go Up

IMPORTS

Distributions | ML_Core | python |

DESCRIPTIONS

ATTRIBUTE Check_Dist

Check_Dist

validation

Go Up

Table of Contents

BinomialRegression.ecl
discrete_GermanDS.ecl
IrisDS.ecl
unit_test_dimm.ecl

validation/

BinomialRegression

Go Up

IMPORTS

validation | ML_Core.Types | ML_Core | Types |

DESCRIPTIONS

ATTRIBUTE BinomialRegression

BinomialRegression

validation/

${\bf discrete_GermanDS}$

Go U	$_{ m Jp}$
<u>IM</u>	<u>PORTS</u>
ML	_Core.Types
DE	SCRIPTIONS
MC	DDULE discrete_GermanDS
C	liscrete_GermanDS
Chil	ldren
1.	content
A T	
ΑI	TRIBUTE content
disc	rete_GermanDS \
C	content

$\frac{\mathrm{validation}}{\mathrm{IrisDS}}$

 ${\rm Go}\ {\rm Up}$

IMPORTS

ML_Core | ML_Core.Types |

DESCRIPTIONS

ATTRIBUTE irisDS

irisDS

$\frac{\mathrm{validation}/}{\mathrm{unit_test_dimm}}$

Go Up

IMPORTS

std.BLAS.Types \mid

DESCRIPTIONS

ATTRIBUTE unit_test_dimm

unit_test_dimm