63
DBMS_DATA_MINING_TRANSFORM

DBMS_DATA_MINING_TRANSFORM implements a set of transformations that are commonly used in machine learning.

This chapter contains the following topics:

> **See Also:**
>
> - DBMS_DATA_MINING
> - *Oracle Machine Learning for SQL User's Guide*

DBMS_DATA_MINING_TRANSFORM Overview

A transformation is a SQL expression that modifies the data in one or more columns.

Data must typically undergo certain transformations before it can be used to build a machine learning model. Many machine learning algorithms have specific transformation requirements.

Data that will be scored must be transformed in the same way as the data that was used to create (train) the model.

External or Embedded Transformations

DBMS_DATA_MINING_TRANSFORM offers two approaches to implementing transformations. For a given model, you can either:

- Create a list of transformation expressions and pass it to the CREATE_MODEL Procedure

  or

- Create a view that implements the transformations and pass the name of the view to the CREATE_MODEL Procedure

If you create a transformation list and pass it to CREATE_MODEL, the transformation expressions are embedded in the model and automatically implemented whenever the model is applied.

If you create a view, the transformation expressions are external to the model. You will need to re-create the transformations whenever you apply the model.

> **Note:**
>
> Embedded transformations significantly enhance the model's usability while simplifying the process of model management.

Automatic Transformations

Oracle Machine Learning for SQL supports an Automatic Data Preparation (ADP) mode. When ADP is enabled, most algorithm-specific transformations are *automatically* embedded. Any additional transformations must be explicitly provided in an embedded transformation list or in a view.

If ADP is enabled and you create a model with a transformation list, both sets of transformations are embedded. The model will execute the user-specified transformations from the transformation list before executing the automatic transformations specified by ADP.

Within a transformation list, you can selectively disable ADP for individual attributes.

> **See Also:**
>
> "Automatic Data Preparation"
>
> *Oracle Machine Learning for SQL User's Guide* for a more information about ADP
>
> "DBMS_DATA_MINING_TRANSFORM-About Transformation Lists"

Transformations in DBMS_DATA_MINING_TRANSFORM

The transformations supported by `DBMS_DATA_MINING_TRANSFORM` are summarized in this section.

Binning

Binning refers to the mapping of continuous or discrete values to discrete values of reduced cardinality.

- Supervised Binning (Categorical and Numerical)

  Binning is based on intrinsic relationships in the data as determined by a decision tree model.

  See "INSERT_BIN_SUPER Procedure".
- Top-N Frequency Categorical Binning

  Binning is based on the number of cases in each category.

  See "INSERT_BIN_CAT_FREQ Procedure"
- Equi-Width Numerical Binning

  Binning is based on equal-range partitions.

  See "INSERT_BIN_NUM_EQWIDTH Procedure".
- Quantile Numerical Binning

  Binning is based on quantiles computed using the SQL `NTILE` function.

See "INSERT_BIN_NUM_QTILE Procedure".

Linear Normalization

Normalization is the process of scaling continuous values down to a specific range, often between zero and one. Normalization transforms each numerical value by subtracting a number (the **shift**) and dividing the result by another number (the **scale**).

```
x_new = (x_old-shift)/scale
```

- Min-Max Normalization

  Normalization is based on the minimum and maximum with the following shift and scale:

  ```
  shift = min
  scale = max-min
  ```

  See "INSERT_NORM_LIN_MINMAX Procedure".

- Scale Normalization

  Normalization is based on the minimum and maximum with the following shift and scale:

  ```
  shift = 0
  scale = max{abs(max), abs(min)}
  ```

  See "INSERT_NORM_LIN_SCALE Procedure".

- Z-Score Normalization

  Normalization is based on the mean and standard deviation with the following shift and scale:

  ```
  shift = mean
  scale = standard_deviation
  ```

  See "INSERT_NORM_LIN_ZSCORE Procedure".

Outlier Treatment

An outlier is a numerical value that is located far from the rest of the data. Outliers can artificially skew the results of machine learning.

- Winsorizing

  Outliers are replaced with the nearest value that is not an outlier.

  See "INSERT_CLIP_WINSOR_TAIL Procedure"

- Trimming

  Outliers are set to `NULL`.

  See "INSERT_CLIP_TRIM_TAIL Procedure".

Missing Value Treatment

Missing data may indicate sparsity or it may indicate that some values are missing at random. `DBMS_DATA_MINING_TRANSFORM` supports the following transformations for minimizing the effects of missing values:

- Missing numerical values are replaced with the mean.

  See "INSERT_MISS_NUM_MEAN Procedure".

- Missing categorical values are replaced with the mode.

  See "INSERT_MISS_CAT_MODE Procedure".

> **Note:**
>
> Oracle Machine Learning for SQL also has default mechanisms for handling missing data. See *Oracle Machine Learning for SQL User's Guide* for details.

DBMS_DATA_MINING_TRANSFORM Operational Notes

The `DBMS_DATA_MINING_TRANSFORM` package offers a flexible framework for specifying data transformations. If you choose to embed transformations in the model (the preferred method), you create a **transformation list** object and pass it to the CREATE_MODEL Procedure. If you choose to transform the data without embedding, you create a view.

When specified in a transformation list, the transformation expressions are run by the model. When specified in a view, the transformation expressions are run by the view.

Transformation Definitions

Transformation definitions are used to generate the SQL expressions that transform the data. For example, the transformation definitions for normalizing a numeric column are the shift and scale values for that data.

With the `DBMS_DATA_MINING_TRANSFORM` package, you can call procedures to compute the transformation definitions, or you can compute them yourself, or you can do both.

Transformation Definition Tables

`DBMS_DATA_MINING_TRANSFORM` provides **`INSERT`** procedures that compute transformation definitions and insert them in transformation definition tables. You can modify the values in the transformation definition tables or populate them yourself.

`XFORM` routines use populated definition tables to transform data in external views. **`STACK`** routines use populated definition tables to build transformation lists.

To specify transformations based on definition tables, follow these steps:

1. Use **`CREATE`** routines to create transformation definition tables.

   The tables have columns to hold the transformation definitions for a given type of transformation. For example, the CREATE_BIN_NUM Procedure creates a definition table that has a column for storing data values and another column for storing the associated bin identifiers.

2. Use **`INSERT`** routines to compute and insert transformation definitions in the tables.

   Each `INSERT` routine uses a specific technique for computing the transformation definitions. For example, the INSERT_BIN_NUM_EQWIDTH Procedure computes bin boundaries by identifying the minimum and maximum values then setting the bin boundaries at equal intervals.

3. Use **`STACK`** or **`XFORM`** routines to generate transformation expressions based on the information in the definition tables:

   - Use **`STACK`** routines to add the transformation expressions to a transformation list. Pass the transformation list to the CREATE_MODEL Procedure. The transformation expressions will be assembled into one long SQL query and embedded in the model.

- Use **XFORM** routines to execute the transformation expressions within a view. The transformations will be external to the model and will need to be re-created whenever the model is applied to new data.

Transformations Without Definition Tables

STACK routines are not the only method for adding transformation expressions to a transformation list. You can also build a transformation list without using definition tables.

To specify transformations without using definition tables, follow these steps:

1. Write a SQL expression for transforming an attribute.
2. Write a SQL expression for reversing the transformation. (See "Reverse Transformations and Model Transparency" in "DBMS_DATA_MINING_TRANSFORM-About Transformation Lists".)
3. Determine whether or not to disable ADP for the attribute. By default ADP is enabled for the attribute if it is specified for the model. (See "Disabling Automatic Data Preparation" in "DBMS_DATA_MINING_TRANSFORM - About Transformation Lists".)
4. Specify the SQL expressions and ADP instructions in a call to the SET_TRANSFORM Procedure, which adds the information to a transformation list.
5. Repeat steps 1 through 4 for each attribute that you wish to transform.
6. Pass the transformation list to the CREATE_MODEL Procedure. The transformation expressions will be assembled into one long SQL query and embedded in the model.

> **Note:**
>
> SQL expressions that you specify with SET_TRANSFORM must fit within a VARCHAR2. To specify a longer expression, you can use the SET_EXPRESSION Procedure. With SET_EXPRESSION, you can build an expression by appending rows to a VARCHAR2 array.

About Stacking

Transformation lists are built by stacking transformation records. Transformation lists are evaluated from bottom to top. Each transformation expression depends on the result of the transformation expression below it in the stack.

Related Topics

- CREATE_MODEL Procedure
  This procedure creates an Oracle Machine Learning for SQL model with a given machine learning function.
- DBMS_DATA_MINING_TRANSFORM — About Transformation Lists
  The elements of a transformation list are **transformation records**. Each transformation record provides all the information needed by the model for managing the transformation of a single attribute.
- DBMS_DATA_MINING_TRANSFORM — About Stacking and Stack Procedures
  Transformation lists are built by stacking transformation records. Transformation lists are evaluated from bottom to top. Each transformation expression depends on the result of the transformation expression below it in the stack.

- DBMS_DATA_MINING_TRANSFORM — Nested Data Transformations
  The `CREATE` routines create transformation definition tables that include two columns, `col` and `att`, for identifying attributes.

DBMS_DATA_MINING_TRANSFORM — About Transformation Lists

The elements of a transformation list are **transformation records**. Each transformation record provides all the information needed by the model for managing the transformation of a single attribute.

Each transformation record includes the following fields:

- *attribute_name* — Name of the column of data to be transformed
- *attribute_subname* — Name of the nested attribute if *attribute_name* is a nested column, otherwise `NULL`
- *expression* — SQL expression for transforming the attribute
- *reverse_expression* — SQL expression for reversing the transformation
- *attribute_spec* — Identifies special treatment for the attribute during the model build. See Table 63-33 for details.

> **See Also:**
>
> - Table 63-1 for details about the `TRANSFORM_LIST` and `TRANSFORM_REC` object types
> - SET_TRANSFORM Procedure
> - CREATE_MODEL Procedure

Reverse Transformations and Model Transparency

An algorithm manipulates transformed attributes to train and score a model. The transformed attributes, however, may not be meaningful to an end user. For example, if attribute *x* has been transformed into bins 1 — 4, the bin names 1, 2 , 3, and 4 are manipulated by the algorithm, but a user is probably not interested in the model details about bins 1 — 4 or in predicting the numbers 1 — 4.

To return original attribute values in model details and predictions, you can provide a reverse expression in the transformation record for the attribute. For example, if you specify the transformation expression `'log(10, y)'` for attribute *y*, you could specify the reverse transformation expression `'power(10, y)'`.

Reverse transformations enable **model transparency**. They make internal processing transparent to the user.

> **Note:**
>
> `STACK` procedures automatically reverse normalization transformations, but they do not provide a mechanism for reversing binning, clipping, or missing value transformations.
>
> You can use the `DBMS_DATA_MINING.ALTER_REVERSE_EXPRESSION` procedure to specify or update reverse transformations expressions for an existing model.

> **See Also:**
>
> Table 63-1
>
> "ALTER_REVERSE_EXPRESSION Procedure"
>
> "Summary of DBMS_DATA_MINING Subprograms" for links to the model details functions

Disabling Automatic Data Preparation

ADP is controlled by a model-specific setting (`PREP_AUTO`). The `PREP_AUTO` setting affects all model attributes unless you disable it for individual attributes.

If ADP is enabled and you set *`attribute_spec`* to `NOPREP`, only the transformations that you specify for that attribute will be evaluated. If ADP is enabled and you do *not* set *`attribute_spec`* to `NOPREP`, the automatic transformations will be evaluated *after* the transformations that you specify for the attribute.

If ADP is not enabled for the model, the *`attribute_spec`* field of the transformation record is ignored.

> **See Also:**
>
> "Automatic Data Preparation" for information about the `PREP_AUTO` setting

Adding Transformation Records to a Transformation List

A transformation list is a stack of transformation records. When a new transformation record is added, it is appended to the top of the stack. (See "About Stacking" for details.)

When you use `SET_TRANSFORM` to add a transformation record to a transformation list, you can specify values for all the fields in the transformation record.

When you use `STACK` procedures to add transformation records to a transformation list, only the transformation expression field is populated. For normalization transformations, the reverse transformation expression field is also populated.

You can use both `STACK` procedures and `SET_TRANSFORM` to build one transformation list. Each `STACK` procedure call adds transformation records for all the attributes in a specified

transformation definition table. Each `SET_TRANSFORM` call adds a transformation record for a single attribute.

DBMS_DATA_MINING_TRANSFORM — About Stacking and Stack Procedures

Transformation lists are built by stacking transformation records. Transformation lists are evaluated from bottom to top. Each transformation expression depends on the result of the transformation expression below it in the stack.

Stack Procedures

`STACK` procedures create transformation records from the information in transformation definition tables. For example `STACK_BIN_NUM` builds a transformation record for each attribute specified in a definition table for numeric binning. `STACK` procedures stack the transformation records as follows:

- If an attribute is specified in the definition table but not in the transformation list, the `STACK` procedure creates a transformation record, computes the reverse transformation (if possible), inserts the transformation and reverse transformation in the transformation record, and appends the transformation record to the top of the transformation list.
- If an attribute is specified in the transformation list but not in the definition table, the `STACK` procedure takes no action.
- If an attribute is specified in the definition table *and* in the transformation list, the `STACK` procedure stacks the transformation expression from the definition table on top of the transformation expression in the transformation record and updates the reverse transformation. See Table 63-1and Example 63-4.

Example 63-1 Stacking a Clipping Transformation

This example shows how STACK_CLIP Procedure would add transformation records to a transformation list. Note that the clipping transformations are not reversed in `COL1` and `COL2` after stacking (as described in "Reverse Transformations and Model Transparency" in "DBMS_DATA_MINING_TRANSFORM-About Transformation Lists").

Refer to:

- CREATE_CLIP Procedure — Creates the definition table
- INSERT_CLIP_TRIM_TAIL Procedure — Inserts definitions in the table
- INSERT_CLIP_WINSOR_TAIL Procedure — Inserts definitions in the table
- Table 63-1 — Describes the structure of the transformation list (`TRANSFORM_LIST` object)

Assume a clipping definition table populated as follows.

col	att	lcut	lval	rcut	rval
COL1	null	-1.5	-1.5	4.5	4.5
COL2	null	0	0	1	1

Assume the following transformation list before stacking.

```
------------------------
transformation record #1:
------------------------
```

```
     attribute_name       = COL1
     attribute_subname    = null
     expression           = log(10, COL1)
     reverse_expression   = power(10, COL1)
-------------------------
transformation record #2:
-------------------------
     attribute_name       = COL3
     attribute_subname    = null
     expression           = ln(COL3)
     reverse_expression   = exp(COL3)
```

After stacking, the transformation list is as follows.

```
-------------------------
transformation record #1:
-------------------------
     attribute_name       = COL1
     attribute_subname    = null
     expression           = CASE WHEN log(10, COL1) < -1.5 THEN -1.5
                                 WHEN log(10, COL1) > 4.5  THEN 4.5
                                 ELSE log(10, COL1)
                            END;
     reverse_expression   = power(10, COL1)
-------------------------
transformation record #2:
-------------------------
     attribute_name       = COL3
     attribute_subname    = null
     expression           = ln(COL3)
     reverse_expression   = exp(COL3)
-------------------------
transformation record #3:
-------------------------
     attribute_name       = COL2
     attribute_subname    = null
     expression           = CASE WHEN COL2 < 0 THEN 0
                                 WHEN COL2 > 1 THEN 1
                                 ELSE COL2
                            END;
     reverse_expression   = null
```

DBMS_DATA_MINING_TRANSFORM — Nested Data Transformations

The `CREATE` routines create transformation definition tables that include two columns, `col` and `att`, for identifying attributes.

The column `col` holds the name of a column in the data table. If the data column is not nested, then `att` is null, and the name of the attribute is *`col`*. If the data column is nested, then `att` holds the name of the nested attribute, and the name of the attribute is *`col.att`*. The `INSERT` and `XFORM` routines ignore the `att` column in the definition tables. Neither the `INSERT` nor the `XFORM` routines support nested data.

Only the `STACK` procedures and `SET_TRANSFORM` support nested data. Nested data transformations are always embedded in the model.

Nested columns in Oracle Machine Learning for SQL can have the following types:

```
DM_NESTED_NUMERICALS
DM_NESTED_CATEGORICALS
```

```
DM_NESTED_BINARY_DOUBLES
DM_NESTED_BINARY_FLOATS
```

> **See Also:**
>
> "Constants"
>
> *Oracle Machine Learning for SQL User's Guide* for details about nested attributes in Oracle Machine Learning for SQL

Specifying Nested Attributes in a Transformation Record

A transformation record (`TRANSFORM_REC`) includes two fields, `attribute_name` and `attribute_subname`, for identifying the attribute. The field `attribute_name` holds the name of a column in the data table. If the data column is not nested, then `attribute_subname` is null, and the name of the attribute is *`attribute_name`*. If the data column is nested, then `attribute_subname` holds the name of the nested attribute, and the name of the attribute is *`attribute_name.attribute_subname`*.

Transforming Individual Nested Attributes

You can specify different transformations for different attributes in a nested column, and you can specify a default transformation for all the remaining attributes in the column. To specify a default nested transformation, specify null in the `attribute_name` field and the name of the nested column in the `attribute_subname` field as shown in Example 63-2. Note that the keyword `VALUE` is used to represent the value of a nested attribute in a transformation expression.

Example 63-2 Transforming a Nested Column

The following statement transforms two of the nested attributes in `COL_N1`. Attribute `ATTR1` is transformed with normalization; Attribute `ATTR2` is set to null, which causes attribute removal transformation (`ATTR2` is not used in training the model). All the remaining attributes in `COL_N1` are divided by 10.

```
DECLARE
  stk dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
  dbms_data_mining_transform.SET_TRANSFORM(
      stk,'COL_N1', 'ATTR1', '(VALUE - (-1.5))/20', 'VALUE *20 + (-1.5)');
  dbms_data_mining_transform.SET_TRANSFORM(
      stk,'COL_N1', 'ATTR2', NULL, NULL);
  dbms_data_mining_transform.SET_TRANSFORM(
      stk, NULL, 'COL_N1', 'VALUE/10', 'VALUE*10');
END;
/
```

The following SQL is generated from this statement.

```
CAST(MULTISET(SELECT DM_NESTED_NUMERICAL(
                                "ATTRIBUTE_NAME",
                                DECODE("ATTRIBUTE_NAME",
                                  'ATTR1', ("VALUE" - (-1.5))/20,
                                  "VALUE"/10))
                        FROM TABLE("COL_N1")
                       WHERE "ATTRIBUTE_NAME" IS NOT IN ('ATTR2'))
             AS DM_NESTED_NUMERICALS)
```

If transformations are not specified for `COL_N1.ATTR1` and `COL_N1.ATTR2`, then the default transformation is used for all the attributes in `COL_N1`, and the resulting SQL does not include a `DECODE`.

```
    CAST(MULTISET(SELECT DM_NESTED_NUMERICAL(
                               "ATTRIBUTE_NAME",
                               "VALUE"/10)
                         FROM TABLE("COL_N1"))
              AS DM_NESTED_NUMERICALS)
```

Since `DECODE` is limited to 256 arguments, multiple `DECODE` functions are nested to support an arbitrary number of individual nested attribute specifications.

Adding a Nested Column

You can specify a transformation that adds a nested column to the data, as shown in Example 63-3.

Example 63-3 Adding a Nested Column to a Transformation List

```
DECLARE
    v_xlst dbms_data_mining_transform.TRANSFORM_LIST;
  BEGIN
    dbms_data_mining_transform.SET_TRANSFORM(v_xlst,
      'YOB_CREDLIM', NULL,
      'dm_nested_numericals(
           dm_nested_numerical(
                ''CUST_YEAR_OF_BIRTH'', cust_year_of_birth),
           dm_nested_numerical(
                ''CUST_CREDIT_LIMIT'', cust_credit_limit))',
        NULL);
    dbms_data_mining_transform.SET_TRANSFORM(
              v_xlst, 'CUST_YEAR_OF_BIRTH', NULL, NULL, NULL);
    dbms_data_mining_transform.SET_TRANSFORM(
              v_xlst, 'CUST_CREDIT_LIMIT', NULL, NULL, NULL);
    dbms_data_mining_transform.XFORM_STACK(
              v_xlst, 'mining_data', 'mining_data_v');
END;
/

set long 2000
SELECT text FROM user_views WHERE view_name IN 'MINING_DATA_V';

TEXT
--------------------------------------------------------------------------------
SELECT "CUST_ID","CUST_POSTAL_CODE",dm_nested_numericals(
        dm_nested_numerical(
           'CUST_YEAR_OF_BIRTH', cust_year_of_birth),
        dm_nested_numerical(
           'CUST_CREDIT_LIMIT', cust_credit_limit)) "YOB_CREDLIM" FROM mining_data

SELECT * FROM mining_data_v WHERE cust_id = 104500;

CUST_ID CUST_POSTAL_CODE YOB_CREDLIM(ATTRIBUTE_NAME, VALUE)
------- ---------------- ------------------------------------------------------------
 104500 68524            DM_NESTED_NUMERICALS(DM_NESTED_NUMERICAL(
                        'CUST_YEAR_OF_BIRTH', 1962),
                         DM_NESTED_NUMERICAL('CUST_CREDIT_LIMIT', 15000))
```

Stacking Nested Transformations

Example 63-4 shows how the STACK_NORM_LIN Procedure would add transformation records for nested column COL_N to a transformation list.

Refer to:

- CREATE_NORM_LIN Procedure — Creates the definition table
- INSERT_NORM_LIN_MINMAX Procedure — Inserts definitions in the table
- INSERT_NORM_LIN_SCALE Procedure — Inserts definitions in the table
- INSERT_NORM_LIN_ZSCORE Procedure — Inserts definitions in the table
- Table 63-1 — Describes the structure of the transformation list

Example 63-4 Stacking a Nested Normalization Transformation

Assume a linear normalization definition table populated as follows.

col	att	shift	scale
COL_N	ATT2	0	20
null	COL_N	0	10

Assume the following transformation list before stacking.

```
-------------------------
transformation record #1:
-------------------------
     attribute_name       = COL_N
     attribute_subname    = ATT1
     expression           = log(10, VALUE)
     reverse_expression   = power(10, VALUE)
-------------------------
transformation record #2:
-------------------------
     attribute_name       = null
     attribute_subname    = COL_N
     expression           = ln(VALUE)
     reverse_expression   = exp(VALUE)
```

After stacking, the transformation list is as follows.

```
-------------------------
transformation record #1:
-------------------------
     attribute_name       = COL_N
     attribute_subname    = ATT1
     expression           = (log(10, VALUE) - 0)/10
     reverse_expression   = power(10, VALUE*10 + 0)
-------------------------
transformation record #2:
-------------------------
     attribute_name       = NULL
     attribute_subname    = COL_N
     expression           = (ln(VALUE)- 0)/10
     reverse_expression   = exp(VALUE *10 + 0)
-------------------------
transformation record #3:
-------------------------
```

```
attribute_name      = COL_N
attribute_subname   = ATT2
expression          = (ln(VALUE) - 0)/20
reverse_expression  = exp(VALUE * 20 + 0)
```

DBMS_DATA_MINING_TRANSFORM Security Model

The DBMS_DATA_MINING_TRANSFORM package is owned by user SYS and is installed as part of database installation. Execution privilege on the package is granted to public. The routines in the package are run with invokers' rights (run with the privileges of the current user).

The DBMS_DATA_MINING_TRANSFORM.INSERT_* procedures have a *data_table_name* parameter that enables the user to provide the input data for transformation purposes. The value of *data_table_name* can be the name of a physical table or a view. The *data_table_name* parameter can also accept an inline query.

> **Note:**
>
> Because an inline query can be used to specify the data for transformation, Oracle strongly recommends that the calling routine perform any necessary SQL injection checks on the input string.

> **See Also:**
>
> "Operational Notes" for a description of the DBMS_DATA_MINING_TRANSFORM.INSERT_* procedures

DBMS_DATA_MINING_TRANSFORM Datatypes

DBMS_DATA_MINING_TRANSFORM defines the datatypes described in the following table.

Table 63-1 Datatypes in DBMS_DATA_MINING_TRANSFORM

List Type	List Elements	Description
COLUMN_LIST	VARRAY(1000) OF varchar2(32)	COLUMN_LIST stores quoted and non-quoted identifiers for column names. COLUMN_LIST is the datatype of the *exclude_list* parameter in the INSERT procedures. See "INSERT_AUTOBIN_NUM_EQWIDTH Procedure" for an example. See *Oracle Database PL/SQL Language Reference* for information about populating VARRAY structures.

Table 63-1 (Cont.) Datatypes in DBMS_DATA_MINING_TRANSFORM

List Type	List Elements	Description
DESCRIBE_LIST	DBMS_SQL.DESC_TAB2 TYPE desc_tab2 IS TABLE OF desc_rec2 INDEX BY BINARY_INTEGER TYPE desc_rec2 IS RECORD (col_type BINARY_INTEGER := 0, col_max_len BINARY_INTEGER := 0, col_name VARCHAR2(32767):= '', col_name_len BINARY_INTEGER := 0, col_schema_name VARCHAR2(32) := '', col_schema_name_len BINARY_INTEGER := 0, col_precision BINARY_INTEGER := 0, col_scale BINARY_INTEGER := 0, col_charsetid BINARY_INTEGER := 0, col_charsetform BINARY_INTEGER := 0, col_null_ok BOOLEAN := TRUE);	DESCRIBE_LIST describes the columns of the data table after the transformation list has been applied. A DESCRIBE_LIST is returned by the DESCRIBE_STACK Procedure. The DESC_TAB2 and DESC_REC2 types are defined in the DBMS_SQL package. See "DESC_REC2 Record Type". The col_type field of DESC_REC2 identifies the datatype of the column. The datatype is expressed as a numeric constant that represents a built-in datatype. For example, a 1 indicates a variable length character string. The codes for Oracle built-in datatypes are listed in *Oracle Database SQL Language Reference*. The codes for the Oracle Machine Learning for SQL nested types are described in "Constants". The col_name field of DESC_REC2 identifies the column name. It may be populated with a column name, an alias, or an expression. If the column name is a SELECT expression, it may be very long. If the expression is longer than 30 bytes, it cannot be used in a view unless it is given an alias.
TRANSFORM_LIST	**TABLE OF transform_rec** **TYPE transform_rec IS RECORD** (attribute_name VARCHAR2(30), attribute_subname VARCHAR2(4000), expression **EXPRESSION_REC**, reverse_expression **EXPRESSION_REC**, attribute_spec VARCHAR2(4000)); **TYPE expression_rec IS RECORD** (lstmt DBMS_SQL.VARCHAR2A, lb BINARY_INTEGER DEFAULT 1, ub BINARY_INTEGER DEFAULT 0); TYPE varchar2a IS TABLE OF VARCHAR2(32767) INDEX BY BINARY_INTEGER;	TRANSFORM_LIST is a list of transformations that can be embedded in a model. A TRANSFORM_LIST is accepted as an argument by the CREATE_MODEL Procedure. Each element in a TRANSFORM_LIST is a TRANSFORM_REC that specifies how to transform a single attribute. The attribute_name is a column name. The attribute_subname is the nested attribute name if the column is nested, otherwise attribute_subname is null. The expression field holds a SQL expression for transforming the attribute. See "About Transformation Lists" for an explanation of reverse expressions. The attribute_spec field can be used to cause the attribute to be handled in a specific way during the model build. See Table 63-33 for details. The expressions in a TRANSFORM_REC have type EXPRESSION_REC. The lstmt field stores a VARCHAR2A, which is a table of VARCHAR2(32767). The VARCHAR2A datatype allows transformation expressions to be very long, as they can be broken up across multiple rows of VARCHAR2. The VARCHAR2A type is defined in the DBMS_SQL package. See "VARCHAR2A Table Type". The ub (upper bound) and lb (lower bound) fields indicate how many rows there are in the VARCHAR2A table. If ub < lb (default) the EXPRESSION_REC is empty; if lb=ub=1 there is one row; if lb=1 and ub=2 there are 2 rows, and so on.

DBMS_DATA_MINING_TRANSFORM Constants

DBMS_DATA_MINING_TRANSFORM defines the constants described in the following table.

Table 63-2 Constants in DBMS_DATA_MINING_TRANSFORM

Constant	Value	Description
NEST_NUM_COL_TYPE	100001	Indicates that an attribute in the transformation list comes from a row in a column of DM_NESTED_NUMERICALS. Nested numerical attributes are defined as follows: attribute_name VARCHAR2(4000) value NUMBER
NEST_CAT_COL_TYPE	100002	Indicates that an attribute in the transformation list comes from a row in a column of DM_NESTED_CATEGORICALS. Nested categorical attributes are defined as follows: attribute_name VARCHAR2(4000) value VARCHAR2(4000)
NEST_BD_COL_TYPE	100003	Indicates that an attribute in the transformation list comes from a row in a column of DM_NESTED_BINARY_DOUBLES. Nested binary double attributes are defined as follows: attribute_name VARCHAR2(4000) value BINARY_DOUBLE
NEST_BF_COL_TYPE	100004	Indicates that an attribute in the transformation list comes from a row in a column of DM_NESTED_BINARY_FLOATS. attribute_name VARCHAR2(4000) value BINARY_FLOAT

See Also:

Oracle Machine Learning for SQL User's Guide for information about nested data in Oracle Machine Learning for SQL

Summary of DBMS_DATA_MINING_TRANSFORM Subprograms

This table lists the DBMS_DATA_MINING_TRANSFORM subprograms in alphabetical order and briefly describes them.

Table 63-3 DBMS_DATA_MINING_TRANSFORM Package Subprograms

Subprogram	Purpose
CREATE_BIN_CAT Procedure	Creates a transformation definition table for categorical binning
CREATE_BIN_NUM Procedure	Creates a transformation definition table for numerical binning
CREATE_CLIP Procedure	Creates a transformation definition table for clipping
CREATE_COL_REM Procedure	Creates a transformation definition table for column removal

Table 63-3 (Cont.) DBMS_DATA_MINING_TRANSFORM Package Subprograms

Subprogram	Purpose
CREATE_MISS_CAT Procedure	Creates a transformation definition table for categorical missing value treatment
CREATE_MISS_NUM Procedure	Creates a transformation definition table for numerical missing values treatment
CREATE_NORM_LIN Procedure	Creates a transformation definition table for linear normalization
DESCRIBE_STACK Procedure	Describes the transformation list
GET_EXPRESSION Function	Returns a `VARCHAR2` chunk from a transformation expression
INSERT_AUTOBIN_NUM_EQWIDTH Procedure	Inserts numeric automatic equi-width binning definitions in a transformation definition table
INSERT_BIN_CAT_FREQ Procedure	Inserts categorical frequency-based binning definitions in a transformation definition table
INSERT_BIN_NUM_EQWIDTH Procedure	Inserts numeric equi-width binning definitions in a transformation definition table
INSERT_BIN_NUM_QTILE Procedure	Inserts numeric quantile binning expressions in a transformation definition table
INSERT_BIN_SUPER Procedure	Inserts supervised binning definitions in numerical and categorical transformation definition tables
INSERT_CLIP_TRIM_TAIL Procedure	Inserts numerical trimming definitions in a transformation definition table
INSERT_CLIP_WINSOR_TAIL Procedure	Inserts numerical winsorizing definitions in a transformation definition table
INSERT_MISS_CAT_MODE Procedure	Inserts categorical missing value treatment definitions in a transformation definition table
INSERT_MISS_NUM_MEAN Procedure	Inserts numerical missing value treatment definitions in a transformation definition table
INSERT_NORM_LIN_MINMAX Procedure	Inserts linear min-max normalization definitions in a transformation definition table
INSERT_NORM_LIN_SCALE Procedure	Inserts linear scale normalization definitions in a transformation definition table
INSERT_NORM_LIN_ZSCORE Procedure	Inserts linear zscore normalization definitions in a transformation definition table
SET_EXPRESSION Procedure	Adds a `VARCHAR2` chunk to an expression
SET_TRANSFORM Procedure	Adds a transformation record to a transformation list
STACK_BIN_CAT Procedure	Adds a categorical binning expression to a transformation list
STACK_BIN_NUM Procedure	Adds a numerical binning expression to a transformation list
STACK_CLIP Procedure	Adds a clipping expression to a transformation list
STACK_COL_REM Procedure	Adds a column removal expression to a transformation list
STACK_MISS_CAT Procedure	Adds a categorical missing value treatment expression to a transformation list
STACK_MISS_NUM Procedure	Adds a numerical missing value treatment expression to a transformation list
STACK_NORM_LIN Procedure	Adds a linear normalization expression to a transformation list
XFORM_BIN_CAT Procedure	Creates a view of the data table with categorical binning transformations

Table 63-3 (Cont.) DBMS_DATA_MINING_TRANSFORM Package Subprograms

Subprogram	Purpose
XFORM_BIN_NUM Procedure	Creates a view of the data table with numerical binning transformations
XFORM_CLIP Procedure	Creates a view of the data table with clipping transformations
XFORM_COL_REM Procedure	Creates a view of the data table with column removal transformations
XFORM_EXPR_NUM Procedure	Creates a view of the data table with the specified numeric transformations
XFORM_EXPR_STR Procedure	Creates a view of the data table with the specified categorical transformations
XFORM_MISS_CAT Procedure	Creates a view of the data table with categorical missing value treatment
XFORM_MISS_NUM Procedure	Creates a view of the data table with numerical missing value treatment
XFORM_NORM_LIN Procedure	Creates a view of the data table with linear normalization transformations
XFORM_STACK Procedure	Creates a view of the transformation list

CREATE_BIN_CAT Procedure

This procedure creates a transformation definition table for categorical binning.

The columns are described in the following table.

Table 63-4 Columns in a Transformation Definition Table for Categorical Binning

Name	Datatype	Description
col	VARCHAR2(30)	Name of a column of categorical data. If the column is not nested, the column name is also the attribute name. For information about attribute names, see *Oracle Machine Learning for SQL User's Guide*.
att	VARCHAR2(4000)	The attribute subname if *col* is a nested column. If *col* is nested, the attribute name is *col.att*. If *col* is not nested, *att* is null.
val	VARCHAR2(4000)	Values of the attribute
bin	VARCHAR2(4000)	Bin assignments for the values

Syntax

```
DBMS_DATA_MINING_TRANSFORM.CREATE_BIN_CAT (
     bin_table_name     IN VARCHAR2,
     bin_schema_name    IN VARCHAR2 DEFAULT NULL );
```

Parameters

Table 63-5 CREATE_BIN_CAT Procedure Parameters

Parameter	Description
bin_table_name	Name of the transformation definition table to be created
bin_schema_name	Schema of *bin_table_name*. If no schema is specified, the current schema is used.

Usage Notes

1. See *Oracle Machine Learning for SQL User's Guide* for details about categorical data.
2. See "Nested Data Transformations" for information about transformation definition tables and nested data.
3. You can use the following procedures to populate the transformation definition table:
   - INSERT_BIN_CAT_FREQ Procedure — frequency-based binning
   - INSERT_BIN_SUPER Procedure — supervised binning

> **See Also:**
>
> "Binning" in DBMS_DATA_MINING_TRANSFORM Overview
>
> "Operational Notes"

Examples

The following statement creates a table called `bin_cat_xtbl` in the current schema. The table has columns that can be populated with bin assignments for categorical attributes.

```
BEGIN
   DBMS_DATA_MINING_TRANSFORM.CREATE_BIN_CAT('bin_cat_xtbl');
END;
/
DESCRIBE bin_cat_xtbl
 Name                                      Null?    Type
 ----------------------------------------- -------- ----------------------------
 COL                                                VARCHAR2(30)
 ATT                                                VARCHAR2(4000)
 VAL                                                VARCHAR2(4000)
 BIN                                                VARCHAR2(4000)
```

CREATE_BIN_NUM Procedure

This procedure creates a transformation definition table for numerical binning.

The columns are described in the following table.

Table 63-6 Columns in a Transformation Definition Table for Numerical Binning

Name	Datatype	Description
col	VARCHAR2(30)	Name of a column of numerical data. If the column is not nested, the column name is also the attribute name. For information about attribute names, see *Oracle Machine Learning for SQL User's Guide*.
att	VARCHAR2(4000)	The attribute subname if *col* is a nested column. If *col* is nested, the attribute name is *col.att*. If *col* is not nested, *att* is null.
val	NUMBER	Values of the attribute
bin	VARCHAR2(4000)	Bin assignments for the values

Syntax

```
DBMS_DATA_MINING_TRANSFORM.CREATE_BIN_NUM (
     bin_table_name    IN VARCHAR2,
     bin_schema_name   IN VARCHAR2 DEFAULT NULL );
```

Parameters

Table 63-7 CREATE_BIN_NUM Procedure Parameters

Parameter	Description
bin_table_name	Name of the transformation definition table to be created
bin_schema_name	Schema of *bin_table_name*. If no schema is specified, the current schema is used.

Usage Notes

1. See *Oracle Machine Learning for SQL User's Guide* for details about numerical data.
2. See "Nested Data Transformations" for information about transformation definition tables and nested data.
3. You can use the following procedures to populate the transformation definition table:
   - INSERT_AUTOBIN_NUM_EQWIDTH Procedure — automatic equi-width binning
   - INSERT_BIN_NUM_EQWIDTH Procedure — user-specified equi-width binning
   - INSERT_BIN_NUM_QTILE Procedure — quantile binning
   - INSERT_BIN_SUPER Procedure — supervised binning

> **See Also:**
>
> "Binning" in DBMS_DATA_MINING_TRANSFORM Overview
>
> "Operational Notes"

Examples

The following statement creates a table called `bin_num_xtbl` in the current schema. The table has columns that can be populated with bin assignments for numerical attributes.

```
BEGIN
  DBMS_DATA_MINING_TRANSFORM.CREATE_BIN_NUM('bin_num_xtbl');
END;
/

DESCRIBE bin_num_xtbl
 Name                                      Null?    Type
 ----------------------------------------- -------- ----------------------------
 COL                                                VARCHAR2(30)
 ATT                                                VARCHAR2(4000)
 VAL                                                NUMBER
 BIN                                                VARCHAR2(4000)
```

CREATE_CLIP Procedure

This procedure creates a transformation definition table for clipping or winsorizing to minimize the effect of outliers.

The columns are described in the following table.

Table 63-8 Columns in a Transformation Definition Table for Clipping or Winsorizing

Name	Datatype	Description
col	VARCHAR2(30)	Name of a column of numerical data. If the column is not nested, the column name is also the attribute name. For information about attribute names, see *Oracle Machine Learning for SQL User's Guide*.
att	VARCHAR2(4000)	The attribute subname if *col* is a nested column of DM_NESTED_NUMERICALS. If *col* is nested, the attribute name is *col.att*. If *col* is not nested, *att* is null.
lcut	NUMBER	The lowest typical value for the attribute. If the attribute values were plotted on an *xy* axis, *lcut* would be the left-most boundary of the range of values considered typical for this attribute. Any values to the left of *lcut* are outliers.
lval	NUMBER	Value assigned to an outlier to the left of *lcut*
rcut	NUMBER	The highest typical value for the attribute If the attribute values were plotted on an *xy* axis, *rcut* would be the right-most boundary of the range of values considered typical for this attribute. Any values to the right of *rcut* are outliers.
rval	NUMBER	Value assigned to an outlier to the right of *rcut*

Syntax

```
DBMS_DATA_MINING_TRANSFORM.CREATE_CLIP (
     clip_table_name    IN VARCHAR2,
     clip_schema_name   IN VARCHAR2 DEFAULT NULL );
```

Parameters

Table 63-9 CREATE_CLIP Procedure Parameters

Parameter	Description
clip_table_name	Name of the transformation definition table to be created
clip_schema_name	Schema of *clip_table_name*. If no schema is specified, the current schema is used.

Usage Notes

1. See *Oracle Machine Learning for SQL User's Guide* for details about numerical data.
2. See "Nested Data Transformations" for information about transformation definition tables and nested data.
3. You can use the following procedures to populate the transformation definition table:
   - INSERT_CLIP_TRIM_TAIL Procedure — replaces outliers with nulls
   - INSERT_CLIP_WINSOR_TAIL Procedure — replaces outliers with an average value

> **See Also:**
>
> "Outlier Treatment" in DBMS_DATA_MINING_TRANSFORM Overview
>
> "Operational Notes"

Examples

The following statement creates a table called clip_xtbl in the current schema. The table has columns that can be populated with clipping instructions for numerical attributes.

```
BEGIN
  DBMS_DATA_MINING_TRANSFORM.CREATE_CLIP('clip_xtbl');
END;
/

DESCRIBE clip_xtbl
 Name                                      Null?    Type
 ----------------------------------------- -------- ----------------------------
 COL                                                VARCHAR2(30)
 ATT                                                VARCHAR2(4000)
 LCUT                                               NUMBER
 LVAL                                               NUMBER
 RCUT                                               NUMBER
 RVAL                                               NUMBER
```

CREATE_COL_REM Procedure

This procedure creates a transformation definition table for removing columns from the data table.

The columns are described in the following table.

Table 63-10 Columns in a Transformation Definition Table for Column Removal

Name	Datatype	Description
col	VARCHAR2(30)	Name of a column of data. If the column is not nested, the column name is also the attribute name. For information about attribute names, see *Oracle Machine Learning for SQL User's Guide*.
att	VARCHAR2(4000)	The attribute subname if *col* is nested (DM_NESTED_NUMERICALS or DM_NESTED_CATEGORICALS). If col is nested, the attribute name is *col.att*. If *col* is not nested, *att* is null.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.CREATE_COL_REM (
      rem_table_name           VARCHAR2,
      rem_schema_name          VARCHAR2 DEFAULT NULL );
```

Parameters

Table 63-11 CREATE_COL_REM Procedure Parameters

Parameter	Description
rem_table_name	Name of the transformation definition table to be created
rem_schema_name	Schema of *rem_table_name*. If no schema is specified, the current schema is used.

Usage Notes

1. See "Nested Data Transformations" for information about transformation definition tables and nested data.
2. See "Operational Notes".

Examples

The following statement creates a table called rem_att_xtbl in the current schema. The table has columns that can be populated with the names of attributes to exclude from the data to be mined.

```
BEGIN
    DBMS_DATA_MINING_TRANSFORM.CREATE_COL_REM ('rem_att_xtbl');
END;
 /
DESCRIBE rem_att_xtbl
 Name                                      Null?    Type
 ----------------------------------------- -------- ----------------------------
 COL                                                VARCHAR2(30)
 ATT                                                VARCHAR2(4000)
```

CREATE_MISS_CAT Procedure

This procedure creates a transformation definition table for replacing categorical missing values.

The columns are described in the following table.

Table 63-12 Columns in a Transformation Definition Table for Categorical Missing Value Treatment

Name	Datatype	Description
col	VARCHAR2(30)	Name of a column of categorical data. If the column is not nested, the column name is also the attribute name. For information about attribute names, see *Oracle Machine Learning for SQL User's Guide*.
att	VARCHAR2(4000)	The attribute subname if *col* is a nested column of DM_NESTED_CATEGORICALS. If *col* is nested, the attribute name is *col.att*. If *col* is not nested, *att* is null.
val	VARCHAR2(4000)	Replacement for missing values in the attribute

Syntax

```
DBMS_DATA_MINING_TRANSFORM.CREATE_MISS_CAT (
     miss_table_name       IN VARCHAR2,
     miss_schema_name      IN VARCHAR2 DEFAULT NULL );
```

Parameters

Table 63-13 CREATE_MISS_CAT Procedure Parameters

Parameter	Description
miss_table_name	Name of the transformation definition table to be created
miss_schema_name	Schema of *miss_table_name*. If no schema is specified, the current schema is used.

Usage Notes

1. See *Oracle Machine Learning for SQL User's Guide* for details about categorical data.
2. See "Nested Data Transformations" for information about transformation definition tables and nested data.
3. You can use the INSERT_MISS_CAT_MODE Procedure to populate the transformation definition table.

> **See Also:**
>
> "Missing Value Treatment" in DBMS_DATA_MINING_TRANSFORM Overview
>
> "Operational Notes"

Examples

The following statement creates a table called `miss_cat_xtbl` in the current schema. The table has columns that can be populated with values for missing data in categorical attributes.

```
BEGIN

  DBMS_DATA_MINING_TRANSFORM.CREATE_MISS_CAT('miss_cat_xtbl');
END;
/

DESCRIBE miss_cat_xtbl
 Name                                      Null?    Type
 ----------------------------------------- -------- ----------------------------
 COL                                                VARCHAR2(30)
 ATT                                                VARCHAR2(4000)
 VAL                                                VARCHAR2(4000)
```

CREATE_MISS_NUM Procedure

This procedure creates a transformation definition table for replacing numerical missing values.

The columns are described in Table 63-14.

Table 63-14 Columns in a Transformation Definition Table for Numerical Missing Value Treatment

Name	Datatype	Description
col	VARCHAR2(30)	Name of a column of numerical data. If the column is not nested, the column name is also the attribute name. For information about attribute names, see *Oracle Machine Learning for SQL User's Guide*.
att	VARCHAR2(4000)	The attribute subname if *col* is a nested column of DM_NESTED_NUMERICALS. If *col* is nested, the attribute name is *col.att*. If *col* is not nested, *att* is null.
val	NUMBER	Replacement for missing values in the attribute

Syntax

```
DBMS_DATA_MINING_TRANSFORM.CREATE_MISS_NUM (
     miss_table_name       IN VARCHAR2,
     miss_schema_name      IN VARCHAR2 DEFAULT NULL );
```

Parameters

Table 63-15 CREATE_MISS_NUM Procedure Parameters

Parameter	Description
miss_table_name	Name of the transformation definition table to be created
miss_schema_name	Schema of *miss_table_name*. If no schema is specified, the current schema is used.

Usage Notes

1. See *Oracle Machine Learning for SQL User's Guide* for details about numerical data.
2. See "Nested Data Transformations" for information about transformation definition tables and nested data.
3. You can use the INSERT_MISS_NUM_MEAN Procedure to populate the transformation definition table.

> **See Also:**
>
> "Missing Value Treatment" in DBMS_DATA_MINING_TRANSFORM Overview
>
> "Operational Notes"

Example

The following statement creates a table called `miss_num_xtbl` in the current schema. The table has columns that can be populated with values for missing data in numerical attributes.

```
BEGIN
    DBMS_DATA_MINING_TRANSFORM.CREATE_MISS_NUM('miss_num_xtbl');
END;
/

DESCRIBE miss_num_xtbl
 Name                                      Null?    Type
 ----------------------------------------- -------- ----------------------------
 COL                                                VARCHAR2(30)
 ATT                                                VARCHAR2(4000)
 VAL                                                NUMBER
```

CREATE_NORM_LIN Procedure

This procedure creates a transformation definition table for linear normalization.

The columns are described in Table 63-16.

Table 63-16 Columns in a Transformation Definition Table for Linear Normalization

Name	Datatype	Description
col	VARCHAR2(30)	Name of a column of numerical data. If the column is not nested, the column name is also the attribute name. For information about attribute names, see *Oracle Machine Learning for SQL User's Guide*.
att	VARCHAR2(4000)	The attribute subname if *col* is a nested column of DM_NESTED_NUMERICALS. If *col* is nested, the attribute name is *col.att*. If *col* is not nested, *att* is null.
shift	NUMBER	A constant to subtract from the attribute values
scale	NUMBER	A constant by which to divide the shifted values

Syntax

```
DBMS_DATA_MINING_TRANSFORM.CREATE_NORM_LIN (
     norm_table_name       IN VARCHAR2,
     norm_schema_name      IN VARCHAR2 DEFAULT NULL );
```

Parameters

Table 63-17 CREATE_NORM_LIN Procedure Parameters

Parameter	Description
norm_table_name	Name of the transformation definition table to be created
norm_schema_name	Schema of *norm_table_name*. If no schema is specified, the current schema is used.

Usage Notes

1. See *Oracle Machine Learning for SQL User's Guide* for details about numerical data.
2. See "Nested Data Transformations" for information about transformation definition tables and nested data.
3. You can use the following procedures to populate the transformation definition table:
   - INSERT_NORM_LIN_MINMAX Procedure — Uses linear min-max normalization
   - INSERT_NORM_LIN_SCALE Procedure — Uses linear scale normalization
   - INSERT_NORM_LIN_ZSCORE Procedure — Uses linear zscore normalization

> **See Also:**
>
> "Linear Normalization" in DBMS_DATA_MINING_TRANSFORM Overview
>
> "Operational Notes"

Examples

The following statement creates a table called norm_xtbl in the current schema. The table has columns that can be populated with shift and scale values for normalizing numerical attributes.

```
BEGIN
     DBMS_DATA_MINING_TRANSFORM.CREATE_NORM_LIN('norm_xtbl');
END;
/

DESCRIBE norm_xtbl
 Name                                      Null?    Type
 ----------------------------------------- -------- ----------------------------
 COL                                                VARCHAR2(30)
 ATT                                                VARCHAR2(4000)
 SHIFT                                              NUMBER
 SCALE                                              NUMBER
```

DESCRIBE_STACK Procedure

This procedure describes the columns of the data table after a list of transformations has been applied.

Only the columns that are specified in the transformation list are transformed. The remaining columns in the data table are included in the output without changes.

To create a view of the data table after the transformations have been applied, use the XFORM_STACK Procedure.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.DESCRIBE_STACK (
     xform_list           IN  TRANSFORM_LIST,
     data_table_name      IN  VARCHAR2,
     describe_list        OUT DESCRIBE_LIST,
     data_schema_name     IN  VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-18 DESCRIBE_STACK Procedure Parameters

Parameter	Description
xform_list	A list of transformations. See Table 63-1 for a description of the TRANSFORM_LIST object type.
data_table_name	Name of the table containing the data to be transformed
describe_list	Descriptions of the columns in the data table after the transformations specified in *xform_list* have been applied. See Table 63-1 for a description of the DESCRIBE_LIST object type.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.

Usage Notes

See "Operational Notes" for information about transformation lists and embedded transformations.

Examples

This example shows the column name and datatype, the column name length, and the column maximum length for the view oml_user.cust_info after the transformation list has been applied. All the transformations are user-specified. The results of DESCRIBE_STACK do not include one of the columns in the original table, because the SET_TRANSFORM procedure sets that column to NULL.

```
CREATE OR REPLACE VIEW cust_info AS
         SELECT a.cust_id, c.country_id, c.cust_year_of_birth,
         CAST(COLLECT(DM_Nested_Numerical(
                  b.prod_name, 1))
                AS DM_Nested_Numericals) custprods
                 FROM sh.sales a, sh.products b, sh.customers c
                  WHERE a.prod_id = b.prod_id AND
                        a.cust_id=c.cust_id and
                        a.cust_id between 100001 AND 105000
         GROUP BY a.cust_id, country_id, cust_year_of_birth;
```

```
describe cust_info
 Name                                      Null?    Type
 ----------------------------------------- -------- ----------------------------
 CUST_ID                                   NOT NULL NUMBER
 COUNTRY_ID                                NOT NULL NUMBER
 CUST_YEAR_OF_BIRTH                        NOT NULL NUMBER(4)
 CUSTPRODS                                          SYS.DM_NESTED_NUMERICALS

DECLARE
  cust_stack   dbms_data_mining_transform.TRANSFORM_LIST;
  cust_cols    dbms_data_mining_transform.DESCRIBE_LIST;
BEGIN
  dbms_data_mining_transform.SET_TRANSFORM (cust_stack,
     'country_id', NULL, 'country_id/10', 'country_id*10');
  dbms_data_mining_transform.SET_TRANSFORM (cust_stack,
      'cust_year_of_birth', NULL, NULL, NULL);
  dbms_data_mining_transform.SET_TRANSFORM (cust_stack,
      'custprods', 'Mouse Pad', 'value*100', 'value/100');
  dbms_data_mining_transform.DESCRIBE_STACK(
       xform_list => cust_stack,
       data_table_name => 'cust_info',
       describe_list => cust_cols);
  dbms_output.put_line('====');
  for i in 1..cust_cols.COUNT loop
    dbms_output.put_line('COLUMN_NAME:     '||cust_cols(i).col_name);
    dbms_output.put_line('COLUMN_TYPE:     '||cust_cols(i).col_type);
    dbms_output.put_line('COLUMN_NAME_LEN: '||cust_cols(i).col_name_len);
    dbms_output.put_line('COLUMN_MAX_LEN:  '||cust_cols(i).col_max_len);
    dbms_output.put_line('====');
  END loop;
END;
/
====
COLUMN_NAME:     CUST_ID
COLUMN_TYPE:     2
COLUMN_NAME_LEN: 7
COLUMN_MAX_LEN:  22
====
COLUMN_NAME:     COUNTRY_ID
COLUMN_TYPE:     2
COLUMN_NAME_LEN: 10
COLUMN_MAX_LEN:  22
====
COLUMN_NAME:     CUSTPRODS
COLUMN_TYPE:     100001
COLUMN_NAME_LEN: 9
COLUMN_MAX_LEN:  40
====
```

GET_EXPRESSION Function

This function returns a row from a `VARCHAR2` array that stores a transformation expression. The array is built by calls to the SET_EXPRESSION Procedure.

The array can be used for specifying SQL expressions that are too long to be used with the SET_TRANSFORM Procedure.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.GET_EXPRESSION (
     expression           IN EXPRESSION_REC,
     chunk_num            IN PLS_INTEGER DEFAULT NULL);
 RETURN VARCHAR2;
```

Parameters

Table 63-19 GET_EXPRESSION Function Parameters

Parameter	Description
expression	An expression record (EXPRESSION_REC) that specifies a transformation expression or a reverse transformation expression for an attribute. Each expression record includes a VARCHAR2 array and index fields for specifying upper and lower boundaries within the array. There are two EXPRESSION_REC fields within a transformation record (TRANSFORM_REC): one for the transformation expression; the other for the reverse transformation expression. See Table 63-1 for a description of the EXPRESSION_REC type.
chunk	A VARCHAR2 chunk (row) to be appended to *expression*.

Usage Notes

1. Chunk numbering starts with one. For chunks outside of the range, the return value is null. When a chunk number is null the whole expression is returned as a string. If the expression is too big, a VALUE_ERROR is raised.
2. See "About Transformation Lists".
3. See "Operational Notes".

Examples

See the example for the SET_EXPRESSION Procedure.

Related Topics

- SET_EXPRESSION Procedure
  This procedure appends a row to a VARCHAR2 array that stores a SQL expression.
- SET_TRANSFORM Procedure
  This procedure appends the transformation instructions for an attribute to a transformation list.

INSERT_AUTOBIN_NUM_EQWIDTH Procedure

This procedure performs numerical binning and inserts the transformation definitions in a transformation definition table. The procedure identifies the minimum and maximum values and computes the bin boundaries at equal intervals.

INSERT_AUTOBIN_NUM_EQWIDTH computes the number of bins separately for each column. If you want to use equi-width binning with the same number of bins for each column, use the INSERT_BIN_NUM_EQWIDTH Procedure.

INSERT_AUTOBIN_NUM_EQWIDTH bins all the NUMBER and FLOAT columns in the data source unless you specify a list of columns to ignore.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.INSERT_AUTOBIN_NUM_EQWIDTH (
     bin_table_name         IN VARCHAR2,
     data_table_name        IN VARCHAR2,
     bin_num                IN PLS_INTEGER DEFAULT 3,
     max_bin_num            IN PLS_INTEGER DEFAULT 100,
     exclude_list           IN COLUMN_LIST DEFAULT NULL,
     round_num              IN PLS_INTEGER DEFAULT 6,
     sample_size            IN PLS_INTEGER DEFAULT 50000,
     bin_schema_name        IN VARCHAR2 DEFAULT NULL,
     data_schema_name       IN VARCHAR2 DEFAULT NULL,
     rem_table_name         IN VARCHAR2 DEFAULT NULL,
     rem_schema_name        IN VARCHAR2 DEFAULT NULL));
```

Parameters

Table 63-20 INSERT_AUTOBIN_NUM_EQWIDTH Procedure Parameters

Parameter	Description
bin_table_name	Name of the transformation definition table for numerical binning. You can use the CREATE_BIN_NUM Procedure to create the definition table. The following columns are required: COL VARCHAR2(30) VAL NUMBER BIN VARCHAR2(4000) CREATE_BIN_NUM creates an additional column, ATT, which may be used for specifying nested attributes. This column is not used by INSERT_AUTOBIN_NUM_EQWIDTH.
data_table_name	Name of the table containing the data to be transformed
bin_num	Minimum number of bins. If *bin_num* is 0 or NULL, it is ignored. The default value of *bin_num* is 3.
max_bin_num	Maximum number of bins. If *max_bin_num* is 0 or NULL, it is ignored. The default value of *max_bin_num* is 100.
exclude_list	List of numerical columns to be excluded from the binning process. If you do not specify *exclude_list*, all numerical columns in the data source are binned. The format of *exclude_list* is: dbms_data_mining_transform.COLUMN_LIST('*col1*','*col2*', ...'*coln*')
round_num	Specifies how to round the number in the VAL column of the transformation definition table. When *round_num* is positive, it specifies the most significant digits to retain. When *round_num* is negative, it specifies the least significant digits to remove. In both cases, the result is rounded to the specified number of digits. See the Usage Notes for an example. The default value of *round_num* is 6.

Table 63-20 (Cont.) INSERT_AUTOBIN_NUM_EQWIDTH Procedure Parameters

Parameter	Description
sample_size	Size of the data sample. If *sample_size* is less than the total number of non-NULL values in the column, then *sample_size* is used instead of the SQL COUNT function in computing the number of bins. If *sample_size* is 0 or NULL, it is ignored. See the Usage Notes. The default value of *sample_size* is 50,000.
bin_schema_name	Schema of *bin_table_name*. If no schema is specified, the current schema is used.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.
rem_table_name	Name of a transformation definition table for column removal. The table must have the columns described in "CREATE_COL_REM Procedure". INSERT_AUTOBIN_NUM_EQWIDTH ignores columns with all nulls or only one unique value. If you specify a value for *rem_table_name*, these columns are removed from the mining data. If you do not specify a value for *rem_table_name*, these unbinned columns remain in the data.
rem_schema_name	Schema of *rem_table_name*. If no schema is specified, the current schema is used.

Usage Notes

1. See *Oracle Machine Learning for SQL User's Guide* for details about numerical data.
2. INSERT_AUTOBIN_NUM_EQWIDTH computes the number of bins for a column based on the number of non-null values (COUNT), the maximum (MAX), the minimum (MIN), the standard deviation (STDDEV), and the constant C=3.49/0.9:

   ```
   N=floor(power(COUNT,1/3)*(max-min)/(c*dev))
   ```

   If the *sample_size* parameter is specified, it is used instead of COUNT.

   See *Oracle Machine Learning for SQL User's Guide* for information about the COUNT, MAX, MIN, STDDEV, FLOOR, and POWER functions.
3. INSERT_AUTOBIN_NUM_EQWIDTH uses absolute values to compute the number of bins. The sign of the parameters *bin_num*, *max_bin_num*, and *sample_size* has no effect on the result.
4. In computing the number of bins, INSERT_AUTOBIN_NUM_EQWIDTH evaluates the following criteria in the following order:
   a. The minimum number of bins (*bin_num*)
   b. The maximum number of bins (*max_bin_num*)
   c. The maximum number of bins for integer columns, calculated as the number of distinct values in the range *max-min+1*.
5. The *round_num* parameter controls the rounding of column values in the transformation definition table, as follows:

   ```
   For a value of 308.162:
   when round_num =  1       result is 300
   when round_num =  2       result is 310
   when round_num =  3       result is 308
   when round_num =  0       result is 308.162
   ```

```
    when round_num = -1       result is 308.16
    when round_num = -2       result is 308.2
```

Examples

In this example, `INSERT_AUTOBIN_NUM_EQWIDTH` computes the bin boundaries for the `cust_year_of_birth` column in `sh.customers` and inserts the transformations in a transformation definition table. The STACK_BIN_NUM Procedure creates a transformation list from the contents of the definition table. The CREATE_MODEL Procedure embeds the transformation list in a new model called `nb_model`.

The transformation and reverse transformation expressions embedded in `nb_model` are returned by the GET_MODEL_TRANSFORMATIONS Function.

```
CREATE OR REPLACE VIEW mining_data AS
       SELECT cust_id, cust_year_of_birth, cust_postal_code
       FROM sh.customers;

DESCRIBE mining_data
 Name                           Null?    Type
 ------------------------------ -------- ----------------------------
 CUST_ID                        NOT NULL NUMBER
 CUST_YEAR_OF_BIRTH             NOT NULL NUMBER(4)
 CUST_POSTAL_CODE               NOT NULL VARCHAR2(10)

BEGIN
  dbms_data_mining_transform.CREATE_BIN_NUM(
     bin_table_name   => 'bin_tbl');
  dbms_data_mining_transform.INSERT_AUTOBIN_NUM_EQWIDTH (
     bin_table_name   => 'bin_tbl',
     data_table_name  => 'mining_data',
     bin_num          => 3,
     max_bin_num      => 5,
     exclude_list     => dbms_data_mining_transform.COLUMN_LIST('cust_id'));
END;
/

set numwidth 4
column val off
SELECT col, val, bin FROM bin_tbl
      ORDER BY val ASC;

COL                         VAL BIN
------------------------- ---- -----
CUST_YEAR_OF_BIRTH        1913
CUST_YEAR_OF_BIRTH        1928 1
CUST_YEAR_OF_BIRTH        1944 2
CUST_YEAR_OF_BIRTH        1959 3
CUST_YEAR_OF_BIRTH        1975 4
CUST_YEAR_OF_BIRTH        1990 5

DECLARE
     year_birth_xform   dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
     dbms_data_mining_transform.STACK_BIN_NUM (
         bin_table_name           => 'bin_tbl',
         xform_list               =>  year_birth_xform);
     dbms_data_mining.CREATE_MODEL(
         model_name              => 'nb_model',
         mining_function         => dbms_data_mining.classification,
         data_table_name         => 'mining_data',
         case_id_column_name     => 'cust_id',
```

```
         target_column_name      => 'cust_postal_code',
         settings_table_name     => null,
         data_schema_name        => null,
         settings_schema_name    => null,
         xform_list              => year_birth_xform);
END;
/

SELECT attribute_name
       FROM TABLE(dbms_data_mining.GET_MODEL_TRANSFORMATIONS('nb_model'));

ATTRIBUTE_NAME
------------------------
CUST_YEAR_OF_BIRTH

SELECT expression
       FROM TABLE(dbms_data_mining.GET_MODEL_TRANSFORMATIONS('nb_model'));

EXPRESSION
--------------------------------------------------------------------------------
CASE WHEN "CUST_YEAR_OF_BIRTH"<1913 THEN NULL WHEN "CUST_YEAR_OF_BIRTH"<=1928.4
 THEN '1' WHEN "CUST_YEAR_OF_BIRTH"<=1943.8 THEN '2' WHEN "CUST_YEAR_OF_BIRTH"
<=1959.2 THEN '3' WHEN "CUST_YEAR_OF_BIRTH"<=1974.6 THEN '4' WHEN
"CUST_YEAR_OF_BIRTH" <=1990 THEN '5' END

SELECT reverse_expression
       FROM TABLE(dbms_data_mining.GET_MODEL_TRANSFORMATIONS('nb_model'));

REVERSE_EXPRESSION
--------------------------------------------------------------------------------
DECODE("CUST_YEAR_OF_BIRTH",'5','(1974.6; 1990]','1','[1913; 1928.4]','2','(1928
.4; 1943.8]','3','(1943.8; 1959.2]','4','(1959.2; 1974.6]',NULL,'( ; 1913), (199
0;  ), NULL')
```

INSERT_BIN_CAT_FREQ Procedure

This procedure performs categorical binning and inserts the transformation definitions in a transformation definition table. The procedure computes the bin boundaries based on frequency.

`INSERT_BIN_CAT_FREQ` bins all the `CHAR` and `VARCHAR2` columns in the data source unless you specify a list of columns to ignore.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.INSERT_BIN_CAT_FREQ (
     bin_table_name       IN VARCHAR2,
     data_table_name      IN VARCHAR2,
     bin_num              IN PLS_INTEGER DEFAULT 9,
     exclude_list         IN COLUMN_LIST DEFAULT NULL,
     default_num          IN PLS_INTEGER DEFAULT 2,
     bin_support          IN NUMBER DEFAULT NULL,
     bin_schema_name      IN VARCHAR2 DEFAULT NULL,
     data_schema_name     IN VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-21 INSERT_BIN_CAT_FREQ Procedure Parameters

Parameter	Description

bin_table_name	Name of the transformation definition table for categorical binning. You can use the CREATE_BIN_CAT Procedure to create the definition table.The following columns are required:
	
	`COL VARCHAR2(30)`
	`VAL VARCHAR2(4000)`
	`BIN VARCHAR2(4000)`
	
	CREATE_BIN_CAT creates an additional column, ATT, which may be used for specifying nested attributes. This column is not used by INSERT_BIN_CAT_FREQ.

data_table_name	Name of the table containing the data to be transformed
bin_num	The number of bins to fill using frequency-based binning The total number of bins will be *bin_num*+1. The additional bin is the default bin. Classes that are not assigned to a frequency-based bin will be assigned to the default bin. The default binning order is from highest to lowest: the most frequently occurring class is assigned to the first bin, the second most frequently occurring class is assigned to the second bin, and so on.You can reverse the binning order by specifying a negative number for *bin_num*. The negative sign causes the binning order to be from lowest to highest. If the total number of distinct values (classes) in the column is less than *bin_num*, then a separate bin will be created for each value and the default bin will be empty. If you specify NULL or 0 for *bin_num*, no binning is performed. The default value of *bin_num* is 9.

exclude_list	List of categorical columns to be excluded from the binning process. If you do not specify *exclude_list*, all categorical columns in the data source are binned.
	
	The format of *exclude_list* is:
	
	`dbms_data_mining_transform.COLUMN_LIST('col1','col2',`
	`...'coln')`

default_num	The number of class occurrences (rows of the same class) required for assignment to the default bin By default, *default_num* is the minimum number of occurrences required for assignment to the default bin. For example, if *default_num* is 3 and a given class occurs only once, it will not be assigned to the default bin. You can change the occurrence requirement from minimum to maximum by specifying a negative number for *default_num*. For example, if *default_num* is -3 and a given class occurs only once, it *will* be assigned to the default bin, but a class that occurs four or more times will not be included. If you specify NULL or 0 for *default_bin*, there are no requirements for assignment to the default bin. The default value of *default_num* is 2.

Table 63-21 (Cont.) INSERT_BIN_CAT_FREQ Procedure Parameters

Parameter	Description
bin_support	The number of class occurrences (rows of the same class) required for assignment to a frequency-based bin. *bin_support* is expressed as a fraction of the total number of rows. By default, *bin_support* is the minimum percentage required for assignment to a frequency-based bin. For example, if there are twenty rows of data and you specify.2 for *bin_support*, then there must be four or more occurrences of a class (.2*20) in order for it to be assigned to a frequency-based bin. You can change *bin_support* from a minimum percentage to a maximum percentage by specifying a negative number for *bin_support*. For example, if there are twenty rows of data and you specify -.2 for *bin_support*, then there must be four or less occurrences of a class in order for it to be assigned to a frequency-based bin. Classes that occur less than a positive *bin_support* or more than a negative *bin_support* will be assigned to the default bin. If you specify NULL or 0 for *bin_support*, then there is no support requirement for frequency-based binning. The default value of *bin_support* is NULL.
bin_schema_name	Schema of *bin_table_name*. If no schema is specified, the current schema is used.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.

Usage Notes

1. See *Oracle Machine Learning for SQL User's Guide* for details about categorical data.
2. If values occur with the same frequency, INSERT_BIN_CAT_FREQ assigns them in descending order when binning is from most to least frequent, or in ascending order when binning is from least to most frequent.

Examples

1. In this example, INSERT_BIN_CAT_FREQ computes the bin boundaries for the cust_postal_code and cust_city columns in sh.customers and inserts the transformations in a transformation definition table. The STACK_BIN_CAT Procedure creates a transformation list from the contents of the definition table, and the CREATE_MODEL Procedure embeds the transformation list in a new model called nb_model.

   The transformation and reverse transformation expressions embedded in nb_model are returned by the GET_MODEL_TRANSFORMATIONS Function.

```
CREATE OR REPLACE VIEW mining_data AS
          SELECT cust_id, cust_year_of_birth, cust_postal_code, cust_city
          FROM sh.customers;

DESCRIBE mining_data
 Name                                      Null?    Type
 ----------------------------------------- -------- ----------------------------
 CUST_ID                                   NOT NULL NUMBER
 CUST_YEAR_OF_BIRTH                        NOT NULL NUMBER(4)
 CUST_POSTAL_CODE                          NOT NULL VARCHAR2(10)
 CUST_CITY                                 NOT NULL VARCHAR2(30)
```

```
BEGIN
    dbms_data_mining_transform.CREATE_BIN_CAT(
       bin_table_name   => 'bin_tbl_1');
    dbms_data_mining_transform.INSERT_BIN_CAT_FREQ (
       bin_table_name   => 'bin_tbl_1',
       data_table_name  => 'mining_data',
       bin_num          => 4);
END;
/

column col format a18
column val format a15
column bin format a10
SELECT col, val, bin
       FROM bin_tbl_1
       ORDER BY col ASC, bin ASC;

COL                VAL             BIN
------------------ --------------- ----------
CUST_CITY          Los Angeles     1
CUST_CITY          Greenwich       2
CUST_CITY          Killarney       3
CUST_CITY          Montara         4
CUST_CITY                          5
CUST_POSTAL_CODE   38082           1
CUST_POSTAL_CODE   63736           2
CUST_POSTAL_CODE   55787           3
CUST_POSTAL_CODE   78558           4
CUST_POSTAL_CODE                   5

DECLARE
      city_xform   dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
      dbms_data_mining_transform.STACK_BIN_CAT (
          bin_table_name          => 'bin_tbl_1',
          xform_list              =>  city_xform);
      dbms_data_mining.CREATE_MODEL(
           model_name             => 'nb_model',
           mining_function        => dbms_data_mining.classification,
           data_table_name        => 'mining_data',
           case_id_column_name    => 'cust_id',
           target_column_name     => 'cust_city',
           settings_table_name    => null,
           data_schema_name       => null,
           settings_schema_name   => null,
           xform_list             => city_xform);
END;
/

SELECT attribute_name
       FROM TABLE(dbms_data_mining.GET_MODEL_TRANSFORMATIONS('nb_model'));

ATTRIBUTE_NAME
--------------------------------------------------------------------------------
CUST_CITY
CUST_POSTAL_CODE

SELECT expression
       FROM TABLE(dbms_data_mining.GET_MODEL_TRANSFORMATIONS('nb_model'));

EXPRESSION
```

```
--------------------------------------------------------------------------------
DECODE("CUST_CITY",'Greenwich','2','Killarney','3','Los Angeles','1',
'Montara','4',NULL,NULL,'5')
DECODE("CUST_POSTAL_CODE",'38082','1','55787','3','63736','2','78558','4',NULL,NULL,'5')

SELECT reverse_expression
       FROM TABLE(dbms_data_mining.GET_MODEL_TRANSFORMATIONS('nb_model'));

REVERSE_EXPRESSION
--------------------------------------------------------------------------------
DECODE("CUST_CITY",'2','''Greenwich''','3','''Killarney''','1',
'''Los Angeles''','4','''Montara''',NULL,'NULL','5','DEFAULT')
DECODE("CUST_POSTAL_CODE",'1','''38082''','3','''55787''','2','''63736''',
'4','''78558''',NULL,'NULL','5','DEFAULT')
```

2. The binning order in example 1 is from most frequent to least frequent. The following example shows reverse order binning (least frequent to most frequent). The binning order is reversed by setting *bin_num* to -4 instead of 4.

```
BEGIN
    dbms_data_mining_transform.CREATE_BIN_CAT(
        bin_table_name   => 'bin_tbl_reverse');
    dbms_data_mining_transform.INSERT_BIN_CAT_FREQ (
        bin_table_name   => 'bin_tbl_reverse',
        data_table_name  => 'mining_data',
        bin_num          => -4);
 END;
 /

column col format a20
SELECT col, val, bin
       FROM bin_tbl_reverse
       ORDER BY col ASC, bin ASC;

COL                  VAL             BIN
-------------------- --------------- ----------
CUST_CITY            Tokyo           1
CUST_CITY            Sliedrecht      2
CUST_CITY            Haarlem         3
CUST_CITY            Diemen          4
CUST_CITY                            5
CUST_POSTAL_CODE     49358           1
CUST_POSTAL_CODE     80563           2
CUST_POSTAL_CODE     74903           3
CUST_POSTAL_CODE     71349           4
CUST_POSTAL_CODE                     5
```

INSERT_BIN_NUM_EQWIDTH Procedure

This procedure performs numerical binning and inserts the transformation definitions in a transformation definition table. The procedure identifies the minimum and maximum values and computes the bin boundaries at equal intervals.

`INSERT_BIN_NUM_EQWIDTH` computes a specified number of bins (*n*) and assigns *(max-min)/n* values to each bin. The number of bins is the same for each column. If you want to use equi-width binning, but you want the number of bins to be calculated on a per-column basis, use the INSERT_AUTOBIN_NUM_EQWIDTH Procedure.

`INSERT_BIN_NUM_EQWIDTH` bins all the `NUMBER` and `FLOAT` columns in the data source unless you specify a list of columns to ignore.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.INSERT_BIN_NUM_EQWIDTH (
     bin_table_name        IN VARCHAR2,
     data_table_name       IN VARCHAR2,
     bin_num               IN PLS_INTEGER DEFAULT 10,
     exclude_list          IN COLUMN_LIST DEFAULT NULL,
     round_num             IN PLS_INTEGER DEFAULT 6,
     bin_schema_name       IN VARCHAR2 DEFAULT NULL,
     data_schema_name      IN VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-22 INSERT_BIN_NUM_EQWIDTH Procedure Parameters

Parameter	Description

bin_table_name	Name of the transformation definition table for numerical binning. You can use the CREATE_BIN_NUM Procedure to create the definition table. The following columns are required:
	`COL VARCHAR2(30)`
	`VAL NUMBER`
	`BIN VARCHAR2(4000)`
	CREATE_BIN_NUM creates an additional column, ATT, which may be used for specifying nested attributes. This column is not used by INSERT_BIN_NUM_EQWIDTH.

data_table_name	Name of the table containing the data to be transformed
bin_num	Number of bins. No binning occurs if *bin_num* is 0 or NULL. The default number of bins is 10.
exclude_list	List of numerical columns to be excluded from the binning process. If you do not specify *exclude_list*, all numerical columns in the data source are binned. The format of *exclude_list* is: `dbms_data_mining_transform.COLUMN_LIST('col1','col2', ...'coln')`
round_num	Specifies how to round the number in the VAL column of the transformation definition table. When *round_num* is positive, it specifies the most significant digits to retain. When *round_num* is negative, it specifies the least significant digits to remove. In both cases, the result is rounded to the specified number of digits. See the Usage Notes for an example. The default value of *round_num* is 6.
bin_schema_name	Schema of *bin_table_name*. If no schema is specified, the current schema is used.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.

Usage Notes

1. See *Oracle Machine Learning for SQL User's Guide* for details about numerical data.

2. The *round_num* parameter controls the rounding of column values in the transformation definition table, as follows:

```
For a value of 308.162:
when round_num =  1       result is 300
when round_num =  2       result is 310
when round_num =  3       result is 308
when round_num =  0       result is 308.162
when round_num = -1       result is 308.16
when round_num = -2       result is 308.2
```

3. INSERT_BIN_NUM_EQWIDTH ignores columns with all NULL values or only one unique value.

Examples

In this example, INSERT_BIN_NUM_EQWIDTH computes the bin boundaries for the affinity_card column in mining_data_build and inserts the transformations in a transformation definition table. The STACK_BIN_NUM Procedure creates a transformation list from the contents of the definition table. The CREATE_MODEL Procedure embeds the transformation list in a new model called glm_model.

The transformation and reverse transformation expressions embedded in glm_model are returned by the GET_MODEL_TRANSFORMATIONS Function.

```
CREATE OR REPLACE VIEW mining_data AS
       SELECT cust_id, cust_income_level, cust_gender, affinity_card
       FROM mining_data_build;

DESCRIBE mining_data
 Name                      Null?    Type
 ------------------------- -------- -----------------
 CUST_ID                   NOT NULL NUMBER
 CUST_INCOME_LEVEL                  VARCHAR2(30)
 CUST_GENDER                        VARCHAR2(1)
 AFFINITY_CARD                      NUMBER(10)

BEGIN
    dbms_data_mining_transform.CREATE_BIN_NUM(
        bin_table_name   => 'bin_tbl');
    dbms_data_mining_transform.INSERT_BIN_NUM_EQWIDTH (
        bin_table_name   => 'bin_tbl',
        data_table_name  => 'mining_data',
        bin_num          => 4,
        exclude_list     => dbms_data_mining_transform.COLUMN_LIST('cust_id'));
END;
/

set numwidth 10
column val off
column col format a20
column bin format a10
SELECT col, val, bin FROM bin_tbl
    ORDER BY val ASC;

COL                         VAL  BIN
-------------------- ----------  ----------
AFFINITY_CARD                 0
AFFINITY_CARD               .25  1
AFFINITY_CARD                .5  2
AFFINITY_CARD               .75  3
AFFINITY_CARD                 1  4
```

```
CREATE TABLE glmsettings(
        setting_name  VARCHAR2(30),
        setting_value VARCHAR2(30));

BEGIN
   INSERT INTO glmsettings (setting_name, setting_value) VALUES
         (dbms_data_mining.algo_name, dbms_data_mining.algo_generalized_linear_model);
   COMMIT;
END;
/

DECLARE
     xforms   dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
     dbms_data_mining_transform.STACK_BIN_NUM (
         bin_table_name           => 'bin_tbl',
         xform_list               =>  xforms,
         literal_flag             =>  TRUE);
     dbms_data_mining.CREATE_MODEL(
         model_name              => 'glm_model',
         mining_function         => dbms_data_mining.regression,
         data_table_name         => 'mining_data',
         case_id_column_name     => 'cust_id',
         target_column_name      => 'affinity_card',
         settings_table_name     => 'glmsettings',
         data_schema_name        => null,
         settings_schema_name    => null,
         xform_list              => xforms);
END;
/

SELECT attribute_name
      FROM TABLE(dbms_data_mining.GET_MODEL_TRANSFORMATIONS('glm_model'));

ATTRIBUTE_NAME
------------------------
AFFINITY_CARD

SELECT expression
       FROM TABLE(dbms_data_mining.GET_MODEL_TRANSFORMATIONS('glm_model'));

EXPRESSION
--------------------------------------------------------------------------------
CASE WHEN "AFFINITY_CARD"<0 THEN NULL WHEN "AFFINITY_CARD"<=.25 THEN 1 WHEN
"AFFINITY_CARD"<=.5 THEN 2 WHEN "AFFINITY_CARD"<=.75 THEN 3 WHEN
"AFFINITY_CARD"<=1 THEN 4 END

SELECT reverse_expression
       FROM TABLE(dbms_data_mining.GET_MODEL_TRANSFORMATIONS('glm_model'));

REVERSE_EXPRESSION
--------------------------------------------------------------------------------
DECODE("AFFINITY_CARD",4,'(.75; 1]',1,'[0; .25]',2,'(.25; .5]',3,'(.5; .75]',
NULL,'( ; 0), (1;  ), NULL')
```

INSERT_BIN_NUM_QTILE Procedure

This procedure performs numerical binning and inserts the transformation definitions in a transformation definition table. The procedure calls the SQL `NTILE` function to order the data and divide it equally into the specified number of bins (quantiles).

`INSERT_BIN_NUM_QTILE` bins all the `NUMBER` and `FLOAT` columns in the data source unless you specify a list of columns to ignore.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.INSERT_BIN_NUM_QTILE (
    bin_table_name       IN VARCHAR2,
    data_table_name      IN VARCHAR2,
    bin_num              IN PLS_INTEGER DEFAULT 10,
    exclude_list         IN COLUMN_LIST DEFAULT NULL,
    bin_schema_name      IN VARCHAR2 DEFAULT NULL,
    data_schema_name     IN VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-23 INSERT_BIN_NUM_QTILE Procedure Parameters

Parameter	Description

`bin_table_name`	Name of the transformation definition table for numerical binning. You can use the CREATE_BIN_NUM Procedure to create the definition table. The following columns are required:
	`COL VARCHAR2(30)`
	`VAL NUMBER`
	`BIN VARCHAR2(4000)`
	`CREATE_BIN_NUM` creates an additional column, `ATT`, which may be used for specifying nested attributes. This column is not used by `INSERT_BIN_NUM_QTILE`.

`data_table_name`	Name of the table containing the data to be transformed
`bin_num`	Number of bins. No binning occurs if *`bin_num`* is `0` or `NULL`. The default number of bins is 10.

`exclude_list`	List of numerical columns to be excluded from the binning process. If you do not specify *`exclude_list`*, all numerical columns in the data source are binned.
	The format of *`exclude_list`* is:
	`dbms_data_mining_transform.COLUMN_LIST('col1','col2',`
	`...'coln')`

`bin_schema_name`	Schema of *`bin_table_name`*. If no schema is specified, the current schema is used.
`data_schema_name`	Schema of *`data_table_name`*. If no schema is specified, the current schema is used.

Usage Notes

1. See *Oracle Machine Learning for SQL User's Guide* for details about numerical data.

2. After dividing the data into quantiles, the `NTILE` function distributes any remainder values one for each quantile, starting with the first. See *Oracle Database SQL Language Reference* for details.
3. Columns with all `NULL` values are ignored by `INSERT_BIN_NUM_QTILE`.

Examples

In this example, `INSERT_BIN_NUM_QTILE` computes the bin boundaries for the `cust_year_of_birth` and `cust_credit_limit` columns in `sh.customers` and inserts the transformations in a transformation definition table. The STACK_BIN_NUM Procedure creates a transformation list from the contents of the definition table.

The SQL expression that computes the transformation is shown in `STACK_VIEW`. The view is for display purposes only; it cannot be used to embed the transformations in a model.

```
CREATE OR REPLACE VIEW mining_data AS
        SELECT cust_id, cust_year_of_birth, cust_credit_limit, cust_city
        FROM sh.customers;

DESCRIBE mining_data
 Name                                    Null?    Type
 ----------------------------------------- -------- ----------------------------
 CUST_ID                                 NOT NULL NUMBER
 CUST_YEAR_OF_BIRTH                      NOT NULL NUMBER(4)
 CUST_CREDIT_LIMIT                       NUMBER
 CUST_CITY                               NOT NULL VARCHAR2(30)

BEGIN
   dbms_data_mining_transform.CREATE_BIN_NUM(
        bin_table_name   => 'bin_tbl');
   dbms_data_mining_transform.INSERT_BIN_NUM_QTILE (
        bin_table_name   => 'bin_tbl',
        data_table_name  => 'mining_data',
        bin_num          => 3,
        exclude_list     => dbms_data_mining_transform.COLUMN_LIST('cust_id'));
END;
/

set numwidth 8
column val off
column col format a20
column bin format a10
SELECT col, val, bin
      FROM bin_tbl
      ORDER BY col ASC, val ASC;

COL                      VAL BIN
-------------------- -------- ----------
CUST_CREDIT_LIMIT        1500
CUST_CREDIT_LIMIT        3000 1
CUST_CREDIT_LIMIT        9000 2
CUST_CREDIT_LIMIT       15000 3
CUST_YEAR_OF_BIRTH       1913
CUST_YEAR_OF_BIRTH       1949 1
CUST_YEAR_OF_BIRTH       1965 2
CUST_YEAR_OF_BIRTH       1990 3

DECLARE
   xforms   dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
   dbms_data_mining_transform.STACK_BIN_NUM (
```

```
        bin_table_name        => 'bin_tbl',
        xform_list            =>  xforms);
   dbms_data_mining_transform.XFORM_STACK (
        xform_list            =>  xforms,
        data_table_name       => 'mining_data',
        xform_view_name       => 'stack_view');
END;
/

set long 3000
SELECT text FROM user_views WHERE view_name in 'STACK_VIEW';

TEXT
--------------------------------------------------------------------------------
SELECT "CUST_ID",CASE WHEN "CUST_YEAR_OF_BIRTH"<1913 THEN NULL WHEN "CUST_YEAR_O
F_BIRTH"<=1949 THEN '1' WHEN "CUST_YEAR_OF_BIRTH"<=1965 THEN '2' WHEN "CUST_YEAR
_OF_BIRTH"<=1990 THEN '3' END "CUST_YEAR_OF_BIRTH",CASE WHEN "CUST_CREDIT_LIMIT"
<1500 THEN NULL WHEN "CUST_CREDIT_LIMIT"<=3000 THEN '1' WHEN "CUST_CREDIT_LIMIT"
<=9000 THEN '2' WHEN "CUST_CREDIT_LIMIT"<=15000 THEN '3' END "CUST_CREDIT_LIMIT"
,"CUST_CITY" FROM mining_data
```

INSERT_BIN_SUPER Procedure

This procedure performs numerical and categorical binning and inserts the transformation definitions in transformation definition tables. The procedure computes bin boundaries based on intrinsic relationships between predictors and a target.

`INSERT_BIN_SUPER` uses an intelligent binning technique known as **supervised binning**. It builds a single-predictor decision tree and derives the bin boundaries from splits within the tree.

`INSERT_BIN_SUPER` bins all the `VARCHAR2`, `CHAR`, `NUMBER`, and `FLOAT` columns in the data source unless you specify a list of columns to ignore.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.INSERT_BIN_SUPER (
    num_table_name        IN VARCHAR2,
    cat_table_name        IN VARCHAR2,
    data_table_name       IN VARCHAR2,
    target_column_name    IN VARCHAR2,
    max_bin_num           IN PLS_INTEGER  DEFAULT 1000,
    exclude_list          IN COLUMN_LIST  DEFAULT NULL,
    num_schema_name       IN VARCHAR2     DEFAULT NULL,
    cat_schema_name       IN VARCHAR2     DEFAULT NULL,
    data_schema_name      IN VARCHAR2     DEFAULT NULL,
    rem_table_name        IN VARCHAR2     DEFAULT NULL,
    rem_schema_name       IN VARCHAR2     DEFAULT NULL);
```

Parameters

Table 63-24 INSERT_BIN_SUPER Procedure Parameters

Parameter	Description

num_table_name	Name of the transformation definition table for numerical binning. You can use the CREATE_BIN_NUM Procedure to create the definition table. The following columns are required:
	`COL VARCHAR2(30)`
	`VAL VNUMBER`
	`BIN VARCHAR2(4000)`
	CREATE_BIN_NUM creates an additional column, ATT, which may be used for specifying nested attributes. This column is not used by INSERT_BIN_SUPER.

cat_table_name	Name of the transformation definition table for categorical binning. You can use the CREATE_BIN_CAT Procedure to create the definition table. The following columns are required:
	`COL VARCHAR2(30)`
	`VAL VARCHAR2(4000)`
	`BIN VARCHAR2(4000)`
	CREATE_BIN_CAT creates an additional column, ATT, which is used for specifying nested attributes. This column is not used by INSERT_BIN_SUPER.

data_table_name	Name of the table containing the data to be transformed
target_column_name	Name of a column to be used as the target for the decision tree models
max_bin_num	The maximum number of bins. The default is 1000.
exclude_list	List of columns to be excluded from the binning process. If you do not specify *exclude_list*, all numerical and categorical columns in the data source are binned. The format of *exclude_list* is: `dbms_data_mining_transform.COLUMN_LIST('col1','col2', ...'coln')`
num_schema_name	Schema of *num_table_name*. If no schema is specified, the current schema is used.
cat_schema_name	Schema of *cat_table_name*. If no schema is specified, the current schema is used.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.
rem_table_name	Name of a column removal definition table. The table must have the columns described in "CREATE_COL_REM Procedure". You can use CREATE_COL_REM to create the table. See Usage Notes.
rem_schema_name	Schema of *rem_table_name*. If no schema is specified, the current schema is used.

Usage Notes

1. See *Oracle Machine Learning for SQL User's Guide* for details about numerical and categorical data.

2. Columns that have no significant splits are not binned. You can remove the unbinned columns from the mining data by specifying a column removal definition table. If you do not specify a column removal definition table, the unbinned columns remain in the mining data.
3. See *Oracle Machine Learning for SQL Concepts* to learn more about decision trees in Oracle Machine Learning for SQL

Examples

In this example, `INSERT_BIN_SUPER` computes the bin boundaries for predictors of `cust_credit_limit` and inserts the transformations in transformation definition tables. One predictor is numerical, the other is categorical. (`INSERT_BIN_SUPER` determines that the `cust_postal_code` column is not a significant predictor.) `STACK` procedures create transformation lists from the contents of the definition tables.

The SQL expressions that compute the transformations are shown in the views `MINING_DATA_STACK_NUM` and `MINING_DATA_STACK_CAT`. The views are for display purposes only; they cannot be used to embed the transformations in a model.

```
CREATE OR REPLACE VIEW mining_data AS
     SELECT cust_id, cust_year_of_birth, cust_marital_status,
            cust_postal_code, cust_credit_limit
     FROM sh.customers;

DESCRIBE mining_data
 Name                           Null?    Type
 ------------------------------ -------- -------------------------------------
 CUST_ID                        NOT NULL NUMBER
 CUST_YEAR_OF_BIRTH             NOT NULL NUMBER(4)
 CUST_MARITAL_STATUS                     VARCHAR2(20)
 CUST_POSTAL_CODE               NOT NULL VARCHAR2(10)
 CUST_CREDIT_LIMIT                       NUMBER

BEGIN
    dbms_data_mining_transform.CREATE_BIN_NUM(
         bin_table_name      => 'bin_num_tbl');
    dbms_data_mining_transform.CREATE_BIN_CAT(
         bin_table_name      => 'bin_cat_tbl');
    dbms_data_mining_transform.CREATE_COL_REM(
         rem_table_name      => 'rem_tbl');
END;
/

BEGIN
   COMMIT;
   dbms_data_mining_transform.INSERT_BIN_SUPER (
      num_table_name      => 'bin_num_tbl',
      cat_table_name      => 'bin_cat_tbl',
      data_table_name     => 'mining_data',
      target_column_name  => 'cust_credit_limit',
      max_bin_num         =>  4,
      exclude_list        =>  dbms_data_mining_transform.COLUMN_LIST('cust_id'),
      num_schema_name     => 'oml_user',
      cat_schema_name     => 'oml_user',
      data_schema_name    => 'oml_user',
      rem_table_name      => 'rem_tbl',
      rem_schema_name     => 'oml_user');
   COMMIT;
END;
/

set numwidth 8
```

```
column val off
SELECT col, val, bin FROM bin_num_tbl
      ORDER BY bin ASC;

COL                       VAL BIN
-------------------- -------- ----------
CUST_YEAR_OF_BIRTH     1923.5 1
CUST_YEAR_OF_BIRTH     1923.5 1
CUST_YEAR_OF_BIRTH     1945.5 2
CUST_YEAR_OF_BIRTH     1980.5 3
CUST_YEAR_OF_BIRTH            4

column val on
column val format a20
SELECT col, val, bin FROM bin_cat_tbl
      ORDER BY bin ASC;

COL                  VAL                  BIN
-------------------- -------------------- ----------
CUST_MARITAL_STATUS  married              1
CUST_MARITAL_STATUS  single               2
CUST_MARITAL_STATUS  Mar-AF               3
CUST_MARITAL_STATUS  Mabsent              3
CUST_MARITAL_STATUS  Divorc.              3
CUST_MARITAL_STATUS  Married              3
CUST_MARITAL_STATUS  Widowed              3
CUST_MARITAL_STATUS  NeverM               3
CUST_MARITAL_STATUS  Separ.               3
CUST_MARITAL_STATUS  divorced             4
CUST_MARITAL_STATUS  widow                4

SELECT col from rem_tbl;

COL
--------------------
CUST_POSTAL_CODE

DECLARE
    xforms_num      dbms_data_mining_transform.TRANSFORM_LIST;
    xforms_cat      dbms_data_mining_transform.TRANSFORM_LIST;
    BEGIN
       dbms_data_mining_transform.STACK_BIN_NUM (
            bin_table_name    => 'bin_num_tbl',
            xform_list        => xforms_num);
       dbms_data_mining_transform.XFORM_STACK (
            xform_list         => xforms_num,
            data_table_name    => 'mining_data',
            xform_view_name    => 'mining_data_stack_num');
       dbms_data_mining_transform.STACK_BIN_CAT (
             bin_table_name    => 'bin_cat_tbl',
             xform_list        => xforms_cat);
       dbms_data_mining_transform.XFORM_STACK (
             xform_list         => xforms_cat,
             data_table_name    => 'mining_data',
             xform_view_name    => 'mining_data_stack_cat');
  END;
 /

set long 3000
SELECT text FROM user_views WHERE view_name IN 'MINING_DATA_STACK_NUM';

TEXT
```

```
--------------------------------------------------------------------------------
SELECT "CUST_ID",CASE WHEN "CUST_YEAR_OF_BIRTH"<1923.5 THEN '1' WHEN "CUST_YEAR_
OF_BIRTH"<=1923.5 THEN '1' WHEN "CUST_YEAR_OF_BIRTH"<=1945.5 THEN '2' WHEN "CUST
_YEAR_OF_BIRTH"<=1980.5 THEN '3' WHEN "CUST_YEAR_OF_BIRTH" IS NOT NULL THEN '4'
END "CUST_YEAR_OF_BIRTH","CUST_MARITAL_STATUS","CUST_POSTAL_CODE","CUST_CREDIT_L
IMIT" FROM mining_data


SELECT text FROM user_views WHERE view_name IN 'MINING_DATA_STACK_CAT';

TEXT
--------------------------------------------------------------------------------
SELECT "CUST_ID","CUST_YEAR_OF_BIRTH",DECODE("CUST_MARITAL_STATUS",'Divorc.','3'
,'Mabsent','3','Mar-AF','3','Married','3','NeverM','3','Separ.','3','Widowed','3
','divorced','4','married','1','single','2','widow','4') "CUST_MARITAL_STATUS","
CUST_POSTAL_CODE","CUST_CREDIT_LIMIT" FROM mining_data
```

INSERT_CLIP_TRIM_TAIL Procedure

This procedure replaces numeric outliers with nulls and inserts the transformation definitions in a transformation definition table.

`INSERT_CLIP_TRIM_TAIL` computes the boundaries of the data based on a specified percentage. It removes the values that fall outside the boundaries (tail values) from the data. If you wish to replace the tail values instead of removing them, use the INSERT_CLIP_WINSOR_TAIL Procedure.

`INSERT_CLIP_TRIM_TAIL` clips all the `NUMBER` and `FLOAT` columns in the data source unless you specify a list of columns to ignore.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.INSERT_CLIP_TRIM_TAIL (
    clip_table_name     IN VARCHAR2,
    data_table_name     IN VARCHAR2,
    tail_frac           IN NUMBER DEFAULT 0.025,
    exclude_list        IN COLUMN_LIST DEFAULT NULL,
    clip_schema_name    IN VARCHAR2 DEFAULT NULL,
    data_schema_name    IN VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-25 INSERT_CLIP_TRIM_TAIL Procedure Parameters

Parameter	Description

`clip_table_name`	Name of the transformation definition table for numerical clipping. You can use the CREATE_CLIP Procedure to create the definition table. The following columns are required:
	
	`COL VARCHAR2(30)`
	`LCUT NUMBER`
	`LVAL NUMBER`
	`RCUT NUMBER`
	`RVAL NUMBER`
	
	`CREATE_CLIP` creates an additional column, `ATT`, which may be used for specifying nested attributes. This column is not used by `INSERT_CLIP_TRIM_TAIL`.

`data_table_name`	Name of the table containing the data to be transformed

Table 63-25 (Cont.) INSERT_CLIP_TRIM_TAIL Procedure Parameters

Parameter	Description
tail_frac	The percentage of non-null values to be designated as outliers at each end of the data. For example, if *tail_frac* is .01, then 1% of the data at the low end and 1% of the data at the high end will be treated as outliers. If *tail_frac* is greater than or equal to .5, no clipping occurs. The default value of *tail_frac* is 0.025.
exclude_list	List of numerical columns to be excluded from the clipping process. If you do not specify *exclude_list*, all numerical columns in the data are clipped. The format of *exclude_list* is: `dbms_data_mining_transform.COLUMN_LIST('col1','col2', ...'coln')`
clip_schema_name	Schema of *clip_table_name*. If no schema is specified, the current schema is used.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.

Usage Notes

1. See *Oracle Machine Learning for SQL User's Guide* for details about numerical data.
2. The DBMS_DATA_MINING_TRANSFORM package provides two clipping procedures: INSERT_CLIP_TRIM_TAIL and INSERT_CLIP_WINSOR_TAIL. Both procedures compute the boundaries as follows:
   - Count the number of non-null values, ***n***, and sort them in ascending order
   - Calculate the number of outliers, ***t***, as ***n*tail_frac***
   - Define the lower boundary ***lcut*** as the value at position ***1+floor(t)***
   - Define the upper boundary ***rcut*** as the value at position ***n-floor(t)***

     (The SQL FLOOR function returns the largest integer less than or equal to ***t***.)
   - All values that are <= ***lcut*** or => ***rcut*** are designated as outliers.

   INSERT_CLIP_TRIM_TAIL replaces the outliers with nulls, effectively removing them from the data.

   INSERT_CLIP_WINSOR_TAIL assigns ***lcut*** to the low outliers and ***rcut*** to the high outliers.

Examples

In this example, INSERT_CLIP_TRIM_TAIL trims 10% of the data in two columns (5% from the high end and 5% from the low end) and inserts the transformations in a transformation definition table. The STACK_CLIP Procedure creates a transformation list from the contents of the definition table.

The SQL expression that computes the trimming is shown in the view MINING_DATA_STACK. The view is for display purposes only; it cannot be used to embed the transformations in a model.

```
CREATE OR REPLACE VIEW mining_data AS
       SELECT cust_id, cust_year_of_birth, cust_credit_limit, cust_city
       FROM sh.customers;
```

```
DESCRIBE mining_data
 Name                           Null?    Type
 ------------------------------ -------- -------------------
 CUST_ID                        NOT NULL NUMBER
 CUST_YEAR_OF_BIRTH             NOT NULL NUMBER(4)
 CUST_CREDIT_LIMIT                       NUMBER
 CUST_CITY                      NOT NULL VARCHAR2(30)

BEGIN
   dbms_data_mining_transform.CREATE_CLIP(
      clip_table_name    => 'clip_tbl');
   dbms_data_mining_transform.INSERT_CLIP_TRIM_TAIL(
      clip_table_name    => 'clip_tbl',
      data_table_name    => 'mining_data',
      tail_frac          => 0.05,
      exclude_list       => DBMS_DATA_MINING_TRANSFORM.COLUMN_LIST('cust_id'));
END;
/

SELECT col, lcut, lval, rcut, rval
      FROM clip_tbl
      ORDER BY col ASC;

COL                     LCUT     LVAL     RCUT     RVAL
------------------- -------- -------- -------- --------
CUST_CREDIT_LIMIT       1500            11000
CUST_YEAR_OF_BIRTH      1934             1982

DECLARE
     xforms      dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
     dbms_data_mining_transform.STACK_CLIP (
          clip_table_name    => 'clip_tbl',
          xform_list         => xforms);
     dbms_data_mining_transform.XFORM_STACK (
          xform_list         => xforms,
          data_table_name    => 'mining_data',
          xform_view_name    => 'mining_data_stack');
 END;
 /

set long 3000
SELECT text FROM user_views WHERE view_name IN 'MINING_DATA_STACK';

TEXT
--------------------------------------------------------------------------------
SELECT "CUST_ID",CASE WHEN "CUST_YEAR_OF_BIRTH" < 1934 THEN NULL WHEN "CUST_YEAR
_OF_BIRTH" > 1982 THEN NULL ELSE "CUST_YEAR_OF_BIRTH" END "CUST_YEAR_OF_BIRTH",C
ASE WHEN "CUST_CREDIT_LIMIT" < 1500 THEN NULL WHEN "CUST_CREDIT_LIMIT" > 11000 T
HEN NULL ELSE "CUST_CREDIT_LIMIT" END "CUST_CREDIT_LIMIT","CUST_CITY" FROM minin
g_data
```

INSERT_CLIP_WINSOR_TAIL Procedure

This procedure replaces numeric outliers with the upper or lower boundary values. It inserts the transformation definitions in a transformation definition table.

`INSERT_CLIP_WINSOR_TAIL` computes the boundaries of the data based on a specified percentage. It replaces the values that fall outside the boundaries (tail values) with the related boundary value. If you wish to set tail values to null, use the INSERT_CLIP_TRIM_TAIL Procedure.

INSERT_CLIP_WINSOR_TAIL clips all the NUMBER and FLOAT columns in the data source unless you specify a list of columns to ignore.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.INSERT_CLIP_WINSOR_TAIL (
    clip_table_name    IN VARCHAR2,
    data_table_name    IN VARCHAR2,
    tail_frac          IN NUMBER DEFAULT 0.025,
    exclude_list       IN COLUMN_LIST DEFAULT NULL,
    clip_schema_name   IN VARCHAR2 DEFAULT NULL,
    data_schema_name   IN VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-26 INSERT_CLIP_WINSOR_TAIL Procedure Parameters

Parameter	Description

clip_table_name	Name of the transformation definition table for numerical clipping. You can use the CREATE_CLIP Procedure to create the definition table. The following columns are required:
	`COL VARCHAR2(30)`
	`LCUT NUMBER`
	`LVAL NUMBER`
	`RCUT NUMBER`
	`RVAL NUMBER`
	CREATE_CLIP creates an additional column, ATT, which may be used for specifying nested attributes. This column is not used by INSERT_CLIP_WINSOR_TAIL.

data_table_name	Name of the table containing the data to be transformed
tail_frac	The percentage of non-null values to be designated as outliers at each end of the data. For example, if *tail_frac* is .01, then 1% of the data at the low end and 1% of the data at the high end will be treated as outliers. If *tail_frac* is greater than or equal to .5, no clipping occurs. The default value of *tail_frac* is 0.025.
exclude_list	List of numerical columns to be excluded from the clipping process. If you do not specify *exclude_list*, all numerical columns in the data are clipped. The format of *exclude_list* is: `dbms_data_mining_transform.COLUMN_LIST('col1','col2', ...'coln')`
clip_schema_name	Schema of *clip_table_name*. If no schema is specified, the current schema is used.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.

Usage Notes

1. See *Oracle Machine Learning for SQL User's Guide* for details about numerical data.
2. The DBMS_DATA_MINING_TRANSFORM package provides two clipping procedures: INSERT_CLIP_WINSOR_TAIL and INSERT_CLIP_TRIM_TAIL. Both procedures compute the boundaries as follows:

- Count the number of non-null values, ***n***, and sort them in ascending order
- Calculate the number of outliers, ***t***, as ***n*tail_frac***
- Define the lower boundary ***lcut*** as the value at position ***1+floor(t)***
- Define the upper boundary ***rcut*** as the value at position ***n-floor(t)***

  (The SQL `FLOOR` function returns the largest integer less than or equal to ***t***.)
- All values that are <= ***lcut*** or => ***rcut*** are designated as outliers.

`INSERT_CLIP_WINSOR_TAIL` assigns ***lcut*** to the low outliers and ***rcut*** to the high outliers.

`INSERT_CLIP_TRIM_TAIL` replaces the outliers with nulls, effectively removing them from the data.

Examples

In this example, `INSERT_CLIP_WINSOR_TAIL` winsorizes 10% of the data in two columns (5% from the high end, and 5% from the low end) and inserts the transformations in a transformation definition table. The STACK_CLIP Procedure creates a transformation list from the contents of the definition table.

The SQL expression that computes the transformation is shown in the view `MINING_DATA_STACK`. The view is for display purposes only; it cannot be used to embed the transformations in a model.

```
CREATE OR REPLACE VIEW mining_data AS
        SELECT cust_id, cust_year_of_birth, cust_credit_limit, cust_city
        FROM sh.customers;

describe mining_data
 Name                                    Null?    Type
 ----------------------------------------- -------- -------------
 CUST_ID                                 NOT NULL NUMBER
 CUST_YEAR_OF_BIRTH                      NOT NULL NUMBER(4)
 CUST_CREDIT_LIMIT                                NUMBER
 CUST_CITY                               NOT NULL VARCHAR2(30)

BEGIN
  dbms_data_mining_transform.CREATE_CLIP(
      clip_table_name   => 'clip_tbl');
  dbms_data_mining_transform.INSERT_CLIP_WINSOR_TAIL(
      clip_table_name   => 'clip_tbl',
      data_table_name   => 'mining_data',
      tail_frac         => 0.05,
      exclude_list      => DBMS_DATA_MINING_TRANSFORM.COLUMN_LIST('cust_id'));
END;
/

SELECT col, lcut, lval, rcut, rval FROM clip_tbl
   ORDER BY col ASC;
COL                            LCUT     LVAL     RCUT     RVAL
------------------------------ -------- -------- -------- --------
CUST_CREDIT_LIMIT                 1500     1500    11000    11000
CUST_YEAR_OF_BIRTH                1934     1934     1982     1982

DECLARE
   xforms      dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
   dbms_data_mining_transform.STACK_CLIP (
   clip_table_name    => 'clip_tbl',
   xform_list        => xforms);
```

```
dbms_data_mining_transform.XFORM_STACK (
   xform_list         => xforms,
   data_table_name    => 'mining_data',
   xform_view_name    => 'mining_data_stack');
END;
/

set long 3000
SQL> SELECT text FROM user_views WHERE view_name IN 'MINING_DATA_STACK';

TEXT
--------------------------------------------------------------------------------
SELECT "CUST_ID",CASE WHEN "CUST_YEAR_OF_BIRTH" < 1934 THEN 1934 WHEN "CUST_YEAR
_OF_BIRTH" > 1982 THEN 1982 ELSE "CUST_YEAR_OF_BIRTH" END "CUST_YEAR_OF_BIRTH",C
ASE WHEN "CUST_CREDIT_LIMIT" < 1500 THEN 1500 WHEN "CUST_CREDIT_LIMIT" > 11000 T
HEN 11000 ELSE "CUST_CREDIT_LIMIT" END "CUST_CREDIT_LIMIT","CUST_CITY" FROM mini
ng_data
```

INSERT_MISS_CAT_MODE Procedure

This procedure replaces missing categorical values with the value that occurs most frequently in the column (the mode). It inserts the transformation definitions in a transformation definition table.

`INSERT_MISS_CAT_MODE` replaces missing values in all `VARCHAR2` and `CHAR` columns in the data source unless you specify a list of columns to ignore.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.INSERT_MISS_CAT_MODE (
    miss_table_name    IN VARCHAR2,
    data_table_name    IN VARCHAR2,
    exclude_list       IN COLUMN_LIST DEFAULT NULL,
    miss_schema_name   IN VARCHAR2 DEFAULT NULL,
    data_schema_name   IN VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-27 INSERT_MISS_CAT_MODE Procedure Parameters

Parameter	Description

`miss_table_name`	Name of the transformation definition table for categorical missing value treatment. You can use the CREATE_MISS_CAT Procedure to create the definition table. The following columns are required:
	`COL VARCHAR2(30)`
	`VAL VARCHAR2(4000)`
	`CREATE_MISS_CAT` creates an additional column, `ATT`, which may be used for specifying nested attributes. This column is not used by `INSERT_MISS_CAT_MODE`.

`data_table_name`	Name of the table containing the data to be transformed

Table 63-27 (Cont.) INSERT_MISS_CAT_MODE Procedure Parameters

Parameter	Description
exclude_list	List of categorical columns to be excluded from missing value treatment. If you do not specify *exclude_list*, all categorical columns are transformed. The format of *exclude_list* is: `dbms_data_mining_transform.COLUMN_LIST('col1','col2', ...'coln')`
miss_schema_name	Schema of *miss_table_name*. If no schema is specified, the current schema is used.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.

Usage Notes

1. See *Oracle Machine Learning for SQL User's Guide* for details about categorical data.
2. If you wish to replace categorical missing values with a value other than the mode, you can edit the transformation definition table.

> **See Also:**
>
> *Oracle Machine Learning for SQL User's Guide* for information about default missing value treatment in Oracle Machine Learning for SQL

Example

In this example, `INSERT_MISS_CAT_MODE` computes missing value treatment for `cust_city` and inserts the transformation in a transformation definition table. The STACK_MISS_CAT Procedure creates a transformation list from the contents of the definition table.

The SQL expression that computes the transformation is shown in the view `MINING_DATA_STACK`. The view is for display purposes only; it cannot be used to embed the transformations in a model.

```
CREATE OR REPLACE VIEW mining_data AS
         SELECT cust_id, cust_year_of_birth, cust_city
         FROM sh.customers;

describe mining_data
 Name                           Null?    Type
 ------------------------------ -------- ----------------
 CUST_ID                        NOT NULL NUMBER
 CUST_YEAR_OF_BIRTH             NOT NULL NUMBER(4)
 CUST_CITY                      NOT NULL VARCHAR2(30)

BEGIN
  dbms_data_mining_transform.create_miss_cat(
      miss_table_name   => 'missc_tbl');
  dbms_data_mining_transform.insert_miss_cat_mode(
      miss_table_name   => 'missc_tbl',
      data_table_name   => 'mining_data');
END;
```

```
/

SELECT stats_mode(cust_city) FROM mining_data;

STATS_MODE(CUST_CITY)
------------------------------
Los Angeles

SELECT col, val
    from missc_tbl;

COL                            VAL
------------------------------ ------------------------------
CUST_CITY                      Los Angeles

DECLARE
    xforms      dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
    dbms_data_mining_transform.STACK_MISS_CAT (
        miss_table_name    => 'missc_tbl',
        xform_list        => xforms);
    dbms_data_mining_transform.XFORM_STACK (
         xform_list         => xforms,
         data_table_name    => 'mining_data',
         xform_view_name    => 'mining_data_stack');
END;
/

set long 3000
SELECT text FROM user_views WHERE view_name IN 'MINING_DATA_STACK';

TEXT
--------------------------------------------------------------------------------
SELECT "CUST_ID","CUST_YEAR_OF_BIRTH",NVL("CUST_CITY",'Los Angeles') "CUST_CITY"
 FROM mining_data
```

INSERT_MISS_NUM_MEAN Procedure

This procedure replaces missing numerical values with the average (the mean) and inserts the transformation definitions in a transformation definition table.

`INSERT_MISS_NUM_MEAN` replaces missing values in all `NUMBER` and `FLOAT` columns in the data source unless you specify a list of columns to ignore.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.INSERT_MISS_NUM_MEAN (
    miss_table_name    IN VARCHAR2,
    data_table_name    IN VARCHAR2,
    exclude_list       IN COLUMN_LIST DEFAULT NULL,
    round_num          IN PLS_INTEGER DEFAULT 6,
    miss_schema_name   IN VARCHAR2 DEFAULT NULL,
    data_schema_name   IN VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-28 INSERT_MISS_NUM_MEAN Procedure Parameters

Parameter	Description

miss_table_name	Name of the transformation definition table for numerical missing value treatment. You can use the CREATE_MISS_NUM Procedure to create the definition table.
	The following columns are required by INSERT_MISS_NUM_MEAN:
	`COL VARCHAR2(30)`
	`VAL NUMBER`
	CREATE_MISS_NUM creates an additional column, ATT, which may be used for specifying nested attributes. This column is not used by INSERT_MISS_NUM_MEAN.

data_table_name	Name of the table containing the data to be transformed
exclude_list	List of numerical columns to be excluded from missing value treatment. If you do not specify *exclude_list*, all numerical columns are transformed. The format of *exclude_list* is: `dbms_data_mining_transform.COLUMN_LIST('col1','col2', ...'coln')`
round_num	The number of significant digits to use for the mean. The default number is 6.
miss_schema_name	Schema of *miss_table_name*. If no schema is specified, the current schema is used.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.

Usage Notes

1. See *Oracle Machine Learning for SQL User's Guide* for details about numerical data.
2. If you wish to replace numerical missing values with a value other than the mean, you can edit the transformation definition table.

> **See Also:**
>
> *Oracle Machine Learning for SQL User's Guide* for information about default missing value treatment in Oracle Machine Learning for SQL

Example

In this example, INSERT_MISS_NUM_MEAN computes missing value treatment for cust_year_of_birth and inserts the transformation in a transformation definition table. The STACK_MISS_NUM Procedure creates a transformation list from the contents of the definition table.

The SQL expression that computes the transformation is shown in the view MINING_DATA_STACK. The view is for display purposes only; it cannot be used to embed the transformations in a model.

```
CREATE OR REPLACE VIEW mining_data AS
    SELECT cust_id, cust_year_of_birth, cust_city
    FROM sh.customers;

DESCRIBE mining_data
 Name                                      Null?    Type
 ----------------------------------------- -------- ----------------------
 CUST_ID                                   NOT NULL NUMBER
 CUST_YEAR_OF_BIRTH                        NOT NULL NUMBER(4)
 CUST_CITY                                 NOT NULL VARCHAR2(30)

BEGIN
   dbms_data_mining_transform.create_miss_num(
       miss_table_name   => 'missn_tbl');
   dbms_data_mining_transform.insert_miss_num_mean(
       miss_table_name   => 'missn_tbl',
       data_table_name   => 'mining_data',
       exclude_list      => DBMS_DATA_MINING_TRANSFORM.COLUMN_LIST('cust_id'));
END;
/

set numwidth 4
column val off
SELECT col, val
  FROM missn_tbl;

COL                   VAL
--------------------- ----
CUST_YEAR_OF_BIRTH    1957

SELECT avg(cust_year_of_birth) FROM mining_data;

AVG(CUST_YEAR_OF_BIRTH)
-----------------------
                   1957

DECLARE
    xforms      dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
    dbms_data_mining_transform.STACK_MISS_NUM (
         miss_table_name    => 'missn_tbl',
         xform_list        => xforms);
    dbms_data_mining_transform.XFORM_STACK (
         xform_list         => xforms,
         data_table_name    => 'mining_data',
         xform_view_name    => 'mining_data_stack');
END;
/

set long 3000
SELECT text FROM user_views WHERE view_name IN 'MINING_DATA_STACK';

TEXT
--------------------------------------------------------------------------------
SELECT "CUST_ID",NVL("CUST_YEAR_OF_BIRTH",1957.4) "CUST_YEAR_OF_BIRTH","CUST_CIT
Y" FROM mining_data
```

INSERT_NORM_LIN_MINMAX Procedure

This procedure performs linear normalization and inserts the transformation definitions in a transformation definition table.

INSERT_NORM_LIN_MINMAX computes the minimum and maximum values from the data and sets the value of *shift* and *scale* as follows:

```
shift = min
scale = max - min
```

Normalization is computed as:

```
x_new = (x_old - shift)/scale
```

INSERT_NORM_LIN_MINMAX rounds the value of *scale* to a specified number of significant digits before storing it in the transformation definition table.

INSERT_NORM_LIN_MINMAX normalizes all the NUMBER and FLOAT columns in the data source unless you specify a list of columns to ignore.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.INSERT_NORM_LIN_MINMAX (
     norm_table_name     IN VARCHAR2,
     data_table_name     IN VARCHAR2,
     exclude_list        IN COLUMN_LIST DEFAULT NULL,
     round_num           IN PLS_INTEGER DEFAULT 6,
     norm_schema_name    IN VARCHAR2 DEFAULT NULL,
     data_schema_name    IN VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-29 *INSERT_NORM_LIN_MINMAX Procedure Parameters*

Parameter	Description

norm_table_name	Name of the transformation definition table for linear normalization. You can use the CREATE_NORM_LIN Procedure to create the definition table. The following columns are required:
	`COL VARCHAR2(30)`
	`SHIFT NUMBER`
	`SCALE NUMBER`
	CREATE_NORM_LIN creates an additional column, ATT, which may be used for specifying nested attributes. This column is not used by INSERT_NORM_LIN_MINMAX.

data_table_name	Name of the table containing the data to be transformed
exclude_list	List of numerical columns to be excluded from normalization. If you do not specify *exclude_list*, all numerical columns are transformed. The format of *exclude_list* is: `dbms_data_mining_transform.COLUMN_LIST('col1','col2', ...'coln')`
round_num	The number of significant digits to use for the minimum and maximum. The default number is 6.

Table 63-29 (Cont.) *INSERT_NORM_LIN_MINMAX Procedure Parameters*

Parameter	Description
norm_schema_name	Schema of *norm_table_name*. If no schema is specified, the current schema is used.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.

Usage Notes

See *Oracle Machine Learning for SQL User's Guide* for details about numerical data.

Examples

In this example, INSERT_NORM_LIN_MINMAX normalizes the cust_year_of_birth column and inserts the transformation in a transformation definition table. The STACK_NORM_LIN Procedure creates a transformation list from the contents of the definition table.

The SQL expression that computes the transformation is shown in the view MINING_DATA_STACK. The view is for display purposes only; it cannot be used to embed the transformations in a model.

```
CREATE OR REPLACE VIEW mining_data AS
         SELECT cust_id, cust_gender, cust_year_of_birth
         FROM sh.customers;

describe mining_data
 Name                                      Null?    Type
 ----------------------------------------- -------- ----------------
 CUST_ID                                   NOT NULL NUMBER
 CUST_GENDER                               NOT NULL CHAR(1)
 CUST_YEAR_OF_BIRTH                        NOT NULL NUMBER(4)

BEGIN
       dbms_data_mining_transform.CREATE_NORM_LIN(
         norm_table_name  => 'norm_tbl');
       dbms_data_mining_transform.INSERT_NORM_LIN_MINMAX(
         norm_table_name  => 'norm_tbl',
         data_table_name  => 'mining_data',
         exclude_list     => dbms_data_mining_transform.COLUMN_LIST( 'cust_id'),
         round_num        => 3);
END;
/

SELECT col, shift, scale FROM norm_tbl;

COL                              SHIFT      SCALE
-------------------------------- ---------- ----------
CUST_YEAR_OF_BIRTH                 1910         77

DECLARE
     xforms      dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
     dbms_data_mining_transform.STACK_NORM_LIN (
          norm_table_name    => 'norm_tbl',
          xform_list        => xforms);
     dbms_data_mining_transform.XFORM_STACK (
          xform_list         => xforms,
```

```
          data_table_name    => 'mining_data',
          xform_view_name    => 'mining_data_stack');
END;
/

set long 3000
SELECT text FROM user_views WHERE view_name IN 'MINING_DATA_STACK';

TEXT
--------------------------------------------------------------------------------
SELECT "CUST_ID","CUST_GENDER",("CUST_YEAR_OF_BIRTH"-1910)/77 "CUST_YEAR_OF_BIRT
H" FROM mining_data
```

INSERT_NORM_LIN_SCALE Procedure

This procedure performs linear normalization and inserts the transformation definitions in a transformation definition table.

`INSERT_NORM_LIN_SCALE` computes the minimum and maximum values from the data and sets the value of *`shift`* and *`scale`* as follows:

```
shift = 0
scale = max(abs(max), abs(min))
```

Normalization is computed as:

```
x_new = (x_old)/scale
```

`INSERT_NORM_LIN_SCALE` rounds the value of *`scale`* to a specified number of significant digits before storing it in the transformation definition table.

`INSERT_NORM_LIN_SCALE` normalizes all the `NUMBER` and `FLOAT` columns in the data source unless you specify a list of columns to ignore.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.INSERT_NORM_LIN_SCALE (
     norm_table_name     IN VARCHAR2,
     data_table_name     IN VARCHAR2,
     exclude_list        IN COLUMN_LIST DEFAULT NULL,
     round_num           IN PLS_INTEGER DEFAULT 6,
     norm_schema_name    IN VARCHAR2 DEFAULT NULL,
     data_schema_name    IN VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-30 INSERT_NORM_LIN_SCALE Procedure Parameters

Parameter	Description

norm_table_name	Name of the transformation definition table for linear normalization. You can use the CREATE_NORM_LIN Procedure to create the definition table. The following columns are required:
	`COL VARCHAR2(30)`
	`SHIFT NUMBER`
	`SCALE NUMBER`
	CREATE_NORM_LIN creates an additional column, ATT, which may be used for specifying nested attributes. This column is not used by INSERT_NORM_LIN_SCALE.

data_table_name	Name of the table containing the data to be transformed
exclude_list	List of numerical columns to be excluded from normalization. If you do not specify *exclude_list*, all numerical columns are transformed. The format of *exclude_list* is: `dbms_data_mining_transform.COLUMN_LIST('col1','col2', ...'coln')`
round_num	The number of significant digits to use for *scale*. The default number is 6.
norm_schema_name	Schema of *norm_table_name*. If no schema is specified, the current schema is used.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.

Usage Notes

See *Oracle Machine Learning for SQL User's Guide* for details about numerical data.

Examples

In this example, INSERT_NORM_LIN_SCALE normalizes the cust_year_of_birth column and inserts the transformation in a transformation definition table. The STACK_NORM_LIN Procedure creates a transformation list from the contents of the definition table.

The SQL expression that computes the transformation is shown in the view MINING_DATA_STACK. The view is for display purposes only; it cannot be used to embed the transformations in a model.

```
CREATE OR REPLACE VIEW mining_data AS
      SELECT cust_id, cust_gender, cust_year_of_birth
      FROM sh.customers;

DESCRIBE mining_data
 Name                              Null?    Type
 --------------------------------- -------- ------------------
 CUST_ID                           NOT NULL NUMBER
 CUST_GENDER                       NOT NULL CHAR(1)
 CUST_YEAR_OF_BIRTH                NOT NULL NUMBER(4)

BEGIN
   dbms_data_mining_transform.CREATE_NORM_LIN(
```

```
       norm_table_name  => 'norm_tbl');
       dbms_data_mining_transform.INSERT_NORM_LIN_SCALE(
       norm_table_name  => 'norm_tbl',
       data_table_name  => 'mining_data',
       exclude_list     => dbms_data_mining_transform.COLUMN_LIST( 'cust_id'),
       round_num        => 3);
 END;
 /

SELECT col, shift, scale FROM norm_tbl;

COL                  SHIFT SCALE
-------------------- ----- -----
CUST_YEAR_OF_BIRTH       0  1990

DECLARE
    xforms      dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
    dbms_data_mining_transform.STACK_NORM_LIN (
        norm_table_name    => 'norm_tbl',
        xform_list        => xforms);
    dbms_data_mining_transform.XFORM_STACK (
        xform_list         => xforms,
        data_table_name    => 'mining_data',
        xform_view_name    => 'mining_data_stack');
END;
/

set long 3000
SELECT text FROM user_views WHERE view_name IN 'MINING_DATA_STACK';

TEXT
--------------------------------------------------------------------------------
SELECT "CUST_ID","CUST_GENDER",("CUST_YEAR_OF_BIRTH"-0)/1990 "CUST_YEAR_OF_BIRTH
" FROM mining_data
```

INSERT_NORM_LIN_ZSCORE Procedure

This procedure performs linear normalization and inserts the transformation definitions in a transformation definition table.

`INSERT_NORM_LIN_ZSCORE` computes the mean and the standard deviation from the data and sets the value of *`shift`* and *`scale`* as follows:

```
shift = mean
scale = stddev
```

Normalization is computed as:

```
x_new = (x_old - shift)/scale
```

`INSERT_NORM_LIN_ZSCORE` rounds the value of *`scale`* to a specified number of significant digits before storing it in the transformation definition table.

`INSERT_NORM_LIN_ZSCORE` normalizes all the `NUMBER` and `FLOAT` columns in the data unless you specify a list of columns to ignore.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.INSERT_NORM_LIN_ZSCORE (
     norm_table_name     IN VARCHAR2,
```

```
    data_table_name     IN VARCHAR2,
    exclude_list        IN COLUMN_LIST DEFAULT NULL,
    round_num           IN PLS_INTEGER DEFAULT 6,
    norm_schema_name    IN VARCHAR2 DEFAULT NULL,
    data_schema_name    IN VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-31 INSERT_NORM_LIN_ZSCORE Procedure Parameters

Parameter	Description

norm_table_name	Name of the transformation definition table for linear normalization. You can use the CREATE_NORM_LIN Procedure to create the definition table. The following columns are required:
	`COL VARCHAR2(30)`
	`SHIFT NUMBER`
	`SCALE NUMBER`
	CREATE_NORM_LIN creates an additional column, ATT, which may be used for specifying nested attributes. This column is not used by INSERT_NORM_LIN_ZSCORE.

data_table_name	Name of the table containing the data to be transformed
exclude_list	List of numerical columns to be excluded from normalization. If you do not specify *exclude_list*, all numerical columns are transformed. The format of *exclude_list* is: `dbms_data_mining_transform.COLUMN_LIST('col1','col2', ...'coln')`
round_num	The number of significant digits to use for *scale*. The default number is 6.
norm_schema_name	Schema of *norm_table_name*. If no schema is specified, the current schema is used.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.

Usage Notes

See *Oracle Machine Learning for SQL User's Guide* for details about numerical data.

Examples

In this example, INSERT_NORM_LIN_ZSCORE normalizes the cust_year_of_birth column and inserts the transformation in a transformation definition table. The STACK_NORM_LIN Procedure creates a transformation list from the contents of the definition table.

The SQL expression that computes the transformation is shown in the view MINING_DATA_STACK. The view is for display purposes only; it cannot be used to embed the transformations in a model.

```
CREATE OR REPLACE VIEW mining_data AS
       SELECT cust_id, cust_gender, cust_year_of_birth
       FROM sh.customers;

DESCRIBE mining_data
 Name                              Null?    Type
 --------------------------------- -------- --------------------
 CUST_ID                           NOT NULL NUMBER
```

```
 CUST_GENDER                          NOT NULL CHAR(1)
 CUST_YEAR_OF_BIRTH                   NOT NULL NUMBER(4)

BEGIN
    dbms_data_mining_transform.CREATE_NORM_LIN(
       norm_table_name  => 'norm_tbl');
       dbms_data_mining_transform.INSERT_NORM_LIN_ZSCORE(
       norm_table_name  => 'norm_tbl',
       data_table_name  => 'mining_data',
       exclude_list     => dbms_data_mining_transform.COLUMN_LIST( 'cust_id'),
       round_num        => 3);
END;
/

SELECT col, shift, scale FROM norm_tbl;

COL                  SHIFT SCALE
-------------------- ----- -----
CUST_YEAR_OF_BIRTH    1960    15

DECLARE
    xforms      dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
    dbms_data_mining_transform.STACK_NORM_LIN (
        norm_table_name    => 'norm_tbl',
        xform_list        => xforms);
    dbms_data_mining_transform.XFORM_STACK (
        xform_list         => xforms,
        data_table_name    => 'mining_data',
        xform_view_name    => 'mining_data_stack');
END;
/

set long 3000
SQL> SELECT text FROM user_views WHERE view_name IN 'MINING_DATA_STACK';

TEXT
--------------------------------------------------------------------------------
SELECT "CUST_ID","CUST_GENDER",("CUST_YEAR_OF_BIRTH"-1960)/15 "CUST_YEAR_OF_BIRT
H" FROM mining_data
```

SET_EXPRESSION Procedure

This procedure appends a row to a `VARCHAR2` array that stores a SQL expression.

The array can be used for specifying a transformation expression that is too long to be used with the SET_TRANSFORM Procedure.

The GET_EXPRESSION Function returns a row in the array.

When you use `SET_EXPRESSION` to build a transformation expression, you must build a corresponding reverse transformation expression, create a transformation record, and add the transformation record to a transformation list.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.SET_EXPRESSION (
           expression    IN OUT NOCOPY EXPRESSION_REC,
           chunk                       VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-32 SET_EXPRESSION Procedure Parameters

Parameter	Description
expression	An expression record (EXPRESSION_REC) that specifies a transformation expression or a reverse transformation expression for an attribute. Each expression record includes a VARCHAR2 array and index fields for specifying upper and lower boundaries within the array. There are two EXPRESSION_REC fields within a transformation record (TRANSFORM_REC): one for the transformation expression; the other for the reverse transformation expression. See Table 63-1 for a description of the EXPRESSION_REC type.
chunk	A VARCHAR2 chunk (row) to be appended to *expression*.

Notes

1. You can pass NULL in the *chunk* argument to SET_EXPRESSION to clear the previous chunk. The default value of *chunk* is NULL.
2. See "About Transformation Lists".
3. See "Operational Notes".

Examples

In this example, two calls to SET_EXPRESSION construct a transformation expression and two calls construct the reverse transformation.

> **Note:**
>
> This example is for illustration purposes only. It shows how SET_EXPRESSION appends the text provided in *chunk* to the text that already exists in *expression*. The SET_EXPRESSION procedure is meant for constructing very long transformation expressions that cannot be specified in a VARCHAR2 argument to SET_TRANSFORM.
>
> Similarly while transformation lists are intended for embedding in a model, the transformation list v_xlst is shown in an external view for illustration purposes.

```
CREATE OR REPLACE VIEW mining_data AS
       SELECT cust_id, cust_year_of_birth, cust_postal_code, cust_credit_limit
       FROM sh.customers;

DECLARE
        v_expr dbms_data_mining_transform.EXPRESSION_REC;
        v_rexp dbms_data_mining_transform.EXPRESSION_REC;
        v_xrec dbms_data_mining_transform.TRANSFORM_REC;
        v_xlst dbms_data_mining_transform.TRANSFORM_LIST :=
                                dbms_data_mining_transform.TRANSFORM_LIST(NULL);
BEGIN
    dbms_data_mining_transform.SET_EXPRESSION(
         EXPRESSION  => v_expr,
         CHUNK       => '("CUST_YEAR_OF_BIRTH"-1910)');
    dbms_data_mining_transform.SET_EXPRESSION(
```

```
          EXPRESSION  => v_expr,
          CHUNK       => '/77');
    dbms_data_mining_transform.SET_EXPRESSION(
          EXPRESSION  => v_rexp,
          CHUNK       => '"CUST_YEAR_OF_BIRTH"*77');
    dbms_data_mining_transform.SET_EXPRESSION(
          EXPRESSION  => v_rexp,
          CHUNK       => '+1910');

    v_xrec := null;
    v_xrec.attribute_name := 'CUST_YEAR_OF_BIRTH';
    v_xrec.expression := v_expr;
    v_xrec.reverse_expression := v_rexp;
    v_xlst.TRIM;
    v_xlst.extend(1);
    v_xlst(1) := v_xrec;

    dbms_data_mining_transform.XFORM_STACK (
        xform_list           =>  v_xlst,
        data_table_name      => 'mining_data',
        xform_view_name      => 'v_xlst_view');

    dbms_output.put_line('====');
    FOR i IN 1..v_xlst.count LOOP
      dbms_output.put_line('ATTR: '||v_xlst(i).attribute_name);
      dbms_output.put_line('SUBN: '||v_xlst(i).attribute_subname);
      FOR j IN v_xlst(i).expression.lb..v_xlst(i).expression.ub LOOP
        dbms_output.put_line('EXPR: '||v_xlst(i).expression.lstmt(j));
      END LOOP;
      FOR j IN v_xlst(i).reverse_expression.lb..
               v_xlst(i).reverse_expression.ub LOOP
        dbms_output.put_line('REXP: '||v_xlst(i).reverse_expression.lstmt(j));
      END LOOP;
      dbms_output.put_line('====');
    END LOOP;
  END;
/
====
ATTR: CUST_YEAR_OF_BIRTH
SUBN:
EXPR: ("CUST_YEAR_OF_BIRTH"-1910)
EXPR: /77
REXP: "CUST_YEAR_OF_BIRTH"*77
REXP: +1910
====
```

SET_TRANSFORM Procedure

This procedure appends the transformation instructions for an attribute to a transformation list.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.SET_TRANSFORM (
          xform_list               IN OUT NOCOPY TRANSFORM_LIST,
          attribute_name           VARCHAR2,
          attribute_subname        VARCHAR2,
          expression               VARCHAR2,
          reverse_expression       VARCHAR2,
          attribute_spec           VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-33 SET_TRANSFORM Procedure Parameters

Parameter	Description
`xform_list`	A transformation list. See Table 63-1for a description of the `TRANSFORM_LIST` object type.
`attribute_name`	Name of the attribute to be transformed
`attribute_subname`	Name of the nested attribute if *`attribute_name`* is a nested column, otherwise `NULL`.
`expression`	A SQL expression that specifies the transformation of the attribute.
`reverse_expression`	A SQL expression that reverses the transformation for readability in model details and in the target of a supervised model (if the attribute is a target)
`attribute_spec`	One or more keywords that identify special treatment for the attribute during model build. Values are: • `NOPREP` — When ADP is on, prevents automatic transformation of the attribute. If ADP is not on, this value has no effect. • `TEXT` — Causes the attribute to be treated as unstructured text data • `FORCE_IN` — Forces the inclusion of the attribute in the model build. Applies only to GLM models with feature selection enabled (`ftr_selection_enable` = yes). Feature selection is disabled by default. If the model is not using GLM with feature selection, this value has no effect. See "Specifying Transformation Instructions for an Attribute" in *Oracle Machine Learning for SQL User's Guide*for more information about `attribute_spec`.

Usage Notes

1. See the following relevant sections in "Operational Notes":
   - About Transformation Lists
   - Nested Data Transformations
2. As shown in the following example, you can eliminate an attribute by specifying a null transformation expression and reverse expression. You can also use the STACK interface to remove a column (CREATE_COL_REM Procedure and STACK_COL_REM Procedure).

STACK_BIN_CAT Procedure

This procedure adds categorical binning transformations to a transformation list.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.STACK_BIN_CAT (
     bin_table_name     IN            VARCHAR2,
     xform_list         IN OUT NOCOPY TRANSFORM_LIST,
     literal_flag       IN            BOOLEAN  DEFAULT FALSE,
     bin_schema_name    IN            VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-34 STACK_BIN_CAT Procedure Parameters

Parameter	Description
bin_table_name	Name of the transformation definition table for categorical binning. You can use the CREATE_BIN_CAT Procedure to create the definition table. The table must be populated with transformation definitions before you call STACK_BIN_CAT. To populate the table, you can use one of the INSERT procedures for categorical binning or you can write your own SQL. See Table 63-4
xform_list	A transformation list. See Table 63-1 for a description of the TRANSFORM_LIST object type.
literal_flag	Indicates whether the values in the bin column in the transformation definition table are valid SQL literals. When *literal_flag* is FALSE (the default), the bin identifiers will be transformed to SQL literals by surrounding them with single quotes. Set *literal_flag* to TRUE if the bin identifiers are numbers that should have a numeric datatype, as is the case for an O-Cluster model. See "INSERT_BIN_NUM_EQWIDTH Procedure" for an example.
bin_schema_name	Schema of *bin_table_name*. If no schema is specified, the current schema is used.

Usage Notes

See "Operational Notes". The following sections are especially relevant:

- "About Transformation Lists"
- "About Stacking"
- "Nested Data Transformations"

Examples

This example shows how a binning transformation for the categorical column cust_postal_code could be added to a stack called mining_data_stack.

> **Note:**
>
> This example invokes the XFORM_STACK Procedure to show how the data is transformed by the stack. XFORM_STACK simply generates an external view of the transformed data. The actual purpose of the STACK procedures is to assemble a list of transformations for embedding in a model. The transformations are passed to CREATE_MODEL in the xform_list parameter. See INSERT_BIN_NUM_EQWIDTH Procedure for an example.

```
CREATE or REPLACE VIEW mining_data AS
    SELECT cust_id, cust_postal_code, cust_credit_limit
       FROM sh.customers
       WHERE cust_id BETWEEN 100050 AND 100100;
BEGIN
   dbms_data_mining_transform.CREATE_BIN_CAT ('bin_cat_tbl');
```

```
   dbms_data_mining_transform.INSERT_BIN_CAT_FREQ (
        bin_table_name   => 'bin_cat_tbl',
        data_table_name  => 'mining_data',
        bin_num          =>  3);
  END;
/
DECLARE
  MINING_DATA_STACK   dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
  dbms_data_mining_transform.STACK_BIN_CAT (
    bin_table_name       => 'bin_cat_tbl',
    xform_list           =>  mining_data_stack);
  dbms_data_mining_transform.XFORM_STACK (
    xform_list           =>  mining_data_stack,
    data_table_name      => 'mining_data',
    xform_view_name      => 'mining_data_stack_view');
  END;
/
-- Before transformation
column cust_postal_code format a16
SELECT * from mining_data
              WHERE cust_id BETWEEN 100050 AND 100053
              ORDER BY cust_id;

  CUST_ID CUST_POSTAL_CODE CUST_CREDIT_LIMIT
---------- ---------------- -----------------
    100050 76486                        1500
    100051 73216                        9000
    100052 69499                        5000
    100053 45704                        7000

-- After transformation
SELECT * FROM mining_data_stack_view
              WHERE cust_id BETWEEN 100050 AND 100053
              ORDER BY cust_id;

  CUST_ID CUST_POSTAL_CODE CUST_CREDIT_LIMIT
---------- ---------------- -----------------
    100050 4                            1500
    100051 1                            9000
    100052 4                            5000
    100053 4                            7000
```

STACK_BIN_NUM Procedure

This procedure adds numerical binning transformations to a transformation list.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.STACK_BIN_NUM (
     bin_table_name     IN             VARCHAR2,
     xform_list         IN OUT  NOCOPY TRANSFORM_LIST,
     literal_flag       IN             BOOLEAN  DEFAULT FALSE,
     bin_schema_name    IN             VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-35 STACK_BIN_NUM Procedure Parameters

Parameter	Description
bin_table_name	Name of the transformation definition table for numerical binning. You can use the CREATE_BIN_NUM Procedure to create the definition table. The table must be populated with transformation definitions before you call STACK_BIN_NUM. To populate the table, you can use one of the INSERT procedures for numerical binning or you can write your own SQL. See Table 63-6.
xform_list	A transformation list. See Table 63-1 for a description of the TRANSFORM_LIST object type.
literal_flag	Indicates whether the values in the bin column in the transformation definition table are valid SQL literals. When *literal_flag* is FALSE (the default), the bin identifiers will be transformed to SQL literals by surrounding them with single quotes. Set *literal_flag* to TRUE if the bin identifiers are numbers that should have a numeric datatype, as is the case for an O-Cluster model. See "INSERT_BIN_NUM_EQWIDTH Procedure" for an example.
bin_schema_name	Schema of *bin_table_name*. If no schema is specified, the current schema is used.

Usage Notes

See "Operational Notes". The following sections are especially relevant:

- "About Transformation Lists"
- "About Stacking"
- "Nested Data Transformations"

Examples

This example shows how a binning transformation for the numerical column cust_credit_limit could be added to a stack called mining_data_stack.

> **Note:**
>
> This example invokes the XFORM_STACK Procedure to show how the data is transformed by the stack. XFORM_STACK simply generates an external view of the transformed data. The actual purpose of the STACK procedures is to assemble a list of transformations for embedding in a model. The transformations are passed to CREATE_MODEL in the xform_list parameter. See INSERT_BIN_NUM_EQWIDTH Procedure for an example.

```
CREATE OR REPLACE VIEW mining_data AS
    SELECT cust_id, cust_postal_code, cust_credit_limit
      FROM sh.customers
      WHERE cust_id BETWEEN 100050 and 100100;
BEGIN
  dbms_data_mining_transform.create_bin_num ('bin_num_tbl');
```

```
  dbms_data_mining_transform.insert_bin_num_qtile (
  bin_table_name    => 'bin_num_tbl',
  data_table_name   => 'mining_data',
  bin_num              => 5,
  exclude_list      =>  dbms_data_mining_transform.COLUMN_LIST('cust_id'));
END;
/
DECLARE
   MINING_DATA_STACK   dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
   dbms_data_mining_transform.STACK_BIN_CAT (
      bin_table_name        => 'bin_num_tbl',
      xform_list            =>  mining_data_stack);
   dbms_data_mining_transform.XFORM_STACK (
      xform_list            =>  mining_data_stack,
      data_table_name   => 'mining_data',
      xform_view_name   => 'mining_data_stack_view');
END;
/
-- Before transformation
SELECT cust_id, cust_postal_code, ROUND(cust_credit_limit) FROM mining_data
   WHERE cust_id BETWEEN 100050 AND 100055
   ORDER BY cust_id;
CUST_ID   CUST_POSTAL_CODE   ROUND(CUST_CREDIT_LIMIT)
-------   -----------------  -------------------------
100050    76486                                 1500
100051    73216                                 9000
100052    69499                                 5000
100053    45704                                 7000
100055    74673                                11000
100055    74673                                11000

-- After transformation
SELECT cust_id, cust_postal_code, ROUND(cust_credit_limit)
   FROM mining_data_stack_view
   WHERE cust_id BETWEEN 100050 AND 100055
   ORDER BY cust_id;
CUST_ID   CUST_POSTAL_CODE   ROUND(CUST_CREDIT_LIMITT)
-------   ----------------   -------------------------
100050    76486
100051    73216                                    2
100052    69499                                    1
100053    45704
100054    88021                                    3
100055    74673                                    3
```

STACK_CLIP Procedure

This procedure adds clipping transformations to a transformation list.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.STACK_CLIP (
          clip_table_name     IN            VARCHAR2,
          xform_list          IN OUT NOCOPY TRANSFORM_LIST,
          clip_schema_name    IN            VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-36 STACK_CLIP Procedure Parameters

Parameter	Description
clip_table_name	Name of the transformation definition table for clipping.You can use the CREATE_CLIP Procedure to create the definition table. The table must be populated with transformation definitions before you call STACK_CLIP. To populate the table, you can use one of the INSERT procedures for clipping or you can write your own SQL. See Table 63-8
xform_list	A transformation list. See Table 63-1 for a description of the TRANSFORM_LIST object type.
clip_schema_name	Schema of *clip_table_name*. If no schema is specified, the current schema is used.

Usage Notes

See DBMS_DATA_MINING_TRANSFORM Operational Notes. The following sections are especially relevant:

- "About Transformation Lists"
- "About Stacking"
- "Nested Data Transformations"

Examples

This example shows how a clipping transformation for the numerical column cust_credit_limit could be added to a stack called mining_data_stack.

> **Note:**
>
> This example invokes the XFORM_STACK Procedure to show how the data is transformed by the stack. XFORM_STACK simply generates an external view of the transformed data. The actual purpose of the STACK procedures is to assemble a list of transformations for embedding in a model. The transformations are passed to CREATE_MODEL in the xform_list parameter. See INSERT_BIN_NUM_EQWIDTH Procedure for an example.

```
CREATE OR REPLACE VIEW mining_data AS
       SELECT cust_id, cust_postal_code, cust_credit_limit
       FROM sh.customers
       WHERE cust_id BETWEEN 100050 AND 100100;
BEGIN
   dbms_data_mining_transform.create_clip ('clip_tbl');
   dbms_data_mining_transform.insert_clip_winsor_tail (
       clip_table_name   => 'clip_tbl',
       data_table_name   => 'mining_data',
       tail_frac         => 0.25,
       exclude_list      => dbms_data_mining_transform.COLUMN_LIST('cust_id'));
END;
/
```

```
DECLARE
      MINING_DATA_STACK   dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
      dbms_data_mining_transform.STACK_CLIP (
         clip_table_name     => 'clip_tbl',
         xform_list          =>  mining_data_stack);
      dbms_data_mining_transform.XFORM_STACK (
          xform_list        =>  mining_data_stack,
          data_table_name   => 'mining_data',
          xform_view_name   => 'mining_data_stack_view');
END;
/
-- Before transformation
SELECT cust_id, cust_postal_code, round(cust_credit_limit)
  FROM mining_data
    WHERE cust_id BETWEEN 100050 AND 100054
    ORDER BY cust_id;

CUST_ID   CUST_POSTAL_CODE   ROUND(CUST_CREDIT_LIMIT)
-------   ----------------   ------------------------
100050    76486                                 1500
100051    73216                                 9000
100052    69499                                 5000
100053    45704                                 7000
100054    88021                                11000

-- After transformation
SELECT cust_id, cust_postal_code, round(cust_credit_limit)
  FROM mining_data_stack_view
    WHERE cust_id BETWEEN 100050 AND 100054
    ORDER BY cust_id;

CUST_ID   CUST_POSTAL_CODE   ROUND(CUST_CREDIT_LIMIT)
-------   ----------------   ------------------------
100050    76486                                 5000
100051    73216                                 9000
100052    69499                                 5000
100053    45704                                 7000
100054    88021                                11000
```

STACK_COL_REM Procedure

This procedure adds column removal transformations to a transformation list.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.STACK_COL_REM (
     rem_table_name     IN            VARCHAR2,
     xform_list         IN OUT NOCOPY TRANSFORM_LIST,
     rem_schema_name    IN            VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-37 STACK_COL_REM Procedure Parameters

Parameter	Description
rem_table_name	Name of the transformation definition table for column removal. You can use the CREATE_COL_REM Procedure to create the definition table. See Table 63-10. The table must be populated with column names before you call STACK_COL_REM. The INSERT_BIN_SUPER Procedure and the INSERT_AUTOBIN_NUM_EQWIDTH Procedure can optionally be used to populate the table. You can also use SQL INSERT statements.
xform_list	A transformation list. See Table 63-1 for a description of the TRANSFORM_LIST object type.
rem_schema_name	Schema of *rem_table_name*. If no schema is specified, the current schema is used.

Usage Notes

See "Operational Notes". The following sections are especially relevant:

- "About Transformation Lists"
- "About Stacking"
- "Nested Data Transformations"

Examples

This example shows how the column cust_credit_limit could be removed in a transformation list called mining_data_stack.

> **Note:**
>
> This example invokes the XFORM_STACK Procedure to show how the data is transformed by the stack. XFORM_STACK simply generates an external view of the transformed data. The actual purpose of the STACK procedures is to assemble a list of transformations for embedding in a model. The transformations are passed to CREATE_MODEL in the xform_list parameter. See INSERT_BIN_NUM_EQWIDTH Procedure for an example.

```
CREATE OR REPLACE VIEW mining_data AS
     SELECT cust_id, country_id, cust_postal_code, cust_credit_limit
        FROM sh.customers;

BEGIN
    dbms_data_mining_transform.create_col_rem ('rem_tbl');
END;
/

INSERT into rem_tbl VALUES (upper('cust_postal_code'), null);

DECLARE
  MINING_DATA_STACK   dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
```

```
     dbms_data_mining_transform.stack_col_rem (
          rem_table_name     => 'rem_tbl',
          xform_list         =>  mining_data_stack);
      dbms_data_mining_transform.XFORM_STACK (
          xform_list         =>  mining_data_stack,
          data_table_name    => 'mining_data',
          xform_view_name    => 'mining_data_stack_view');
END;
/

SELECT *  FROM mining_data
  WHERE cust_id BETWEEN 100050 AND 100051
  ORDER BY cust_id;

CUST_ID   COUNTRY_ID   CUST_POSTAL_CODE   CUST_CREDIT_LIMIT
-------   ----------   ----------------   -----------------
100050         52773   76486                          1500
100051         52790   73216                          9000

SELECT *  FROM mining_data_stack_view
  WHERE cust_id BETWEEN 100050 AND 100051
  ORDER BY cust_id;

CUST_ID   COUNTRY_ID   CUST_CREDIT_LIMIT
-------   ----------   -----------------
100050         52773              1500
100051         52790              9000
```

STACK_MISS_CAT Procedure

This procedure adds categorical missing value transformations to a transformation list.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.STACK_MISS_CAT (
     miss_table_name     IN       VARCHAR2,
     xform_list          IN OUT   NOCOPY TRANSFORM_LIST,
     miss_schema_name    IN       VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-38 STACK_MISS_CAT Procedure Parameters

Parameter	Description
`miss_table_name`	Name of the transformation definition table for categorical missing value treatment. You can use the CREATE_MISS_CAT Procedure to create the definition table. The table must be populated with transformation definitions before you call `STACK_MISS_CAT`. To populate the table, you can use the INSERT_MISS_CAT_MODE Procedure or you can write your own SQL. See Table 63-12.
`xform_list`	A transformation list. See Table 63-1 for a description of the `TRANSFORM_LIST` object type.
`miss_schema_name`	Schema of *`miss_table_name`*. If no schema is specified, the current schema is used.

Usage Notes

See "Operational Notes". The following sections are especially relevant:

- "About Transformation Lists"
- "About Stacking"
- "Nested Data Transformations"

Examples

This example shows how the missing values in the column `cust_marital_status` could be replaced with the mode in a transformation list called `mining_data_stack`.

> **Note:**
>
> This example invokes the XFORM_STACK Procedure to show how the data is transformed by the stack. `XFORM_STACK` simply generates an external view of the transformed data. The actual purpose of the `STACK` procedures is to assemble a list of transformations for embedding in a model. The transformations are passed to `CREATE_MODEL` in the `xform_list` parameter. See INSERT_BIN_NUM_EQWIDTH Procedure for an example.

```
CREATE OR REPLACE VIEW mining_data AS
      SELECT cust_id, country_id, cust_marital_status
         FROM sh.customers
         where cust_id BETWEEN 1 AND 10;

BEGIN
  dbms_data_mining_transform.create_miss_cat ('miss_cat_tbl');
  dbms_data_mining_transform.insert_miss_cat_mode ('miss_cat_tbl', 'mining_data');
END;
/

DECLARE
  MINING_DATA_STACK   dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
     dbms_data_mining_transform.stack_miss_cat (
          miss_table_name   => 'miss_cat_tbl',
          xform_list        => mining_data_stack);
      dbms_data_mining_transform.XFORM_STACK (
          xform_list        =>  mining_data_stack,
          data_table_name   => 'mining_data',
          xform_view_name   => 'mining_data_stack_view');
END;
/
SELECT * FROM mining_data
  ORDER BY cust_id;

CUST_ID   COUNTRY_ID   CUST_MARITAL_STATUS
-------   ----------   --------------------
      1        52789
      2        52778
      3        52770
      4        52770
      5        52789
      6        52769     single
      7        52790     single
      8        52790     married
      9        52770     divorced
     10        52790     widow
```

```
SELECT * FROM mining_data_stack_view
   ORDER By cust_id;

CUST_ID   COUNTRY_ID   CUST_MARITAL_STATUS
-------   ----------  --------------------
      1        52789   single
      2        52778   single
      3        52770   single
      4        52770   single
      5        52789   single
      6        52769   single
      7        52790   single
      8        52790   married
      9        52770   divorced
     10        52790   widow
```

STACK_MISS_NUM Procedure

This procedure adds numeric missing value transformations to a transformation list.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.STACK_MISS_NUM (
     miss_table_name     IN       VARCHAR2,
     xform_list          IN OUT   NOCOPY TRANSFORM_LIST,
     miss_schema_name    IN       VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-39 STACK_MISS_NUM Procedure Parameters

Parameter	Description
`miss_table_name`	Name of the transformation definition table for numerical missing value treatment. You can use the CREATE_MISS_NUM Procedure to create the definition table. The table must be populated with transformation definitions before you call `STACK_MISS_NUM`. To populate the table, you can use the INSERT_MISS_NUM_MEAN Procedure or you can write your own SQL. See Table 63-14.
`xform_list`	A transformation list. See Table 63-1 for a description of the `TRANSFORM_LIST` object type.
`miss_schema_name`	Schema of *`miss_table_name`*. If no schema is specified, the current schema is used.

Usage Notes

See "Operational Notes". The following sections are especially relevant:

- "About Transformation Lists"
- "About Stacking"
- "Nested Data Transformations"

Examples

This example shows how the missing values in the column `cust_credit_limit` could be replaced with the mean in a transformation list called `mining_data_stack`.

> **Note:**
>
> This example invokes the XFORM_STACK Procedure to show how the data is transformed by the stack. `XFORM_STACK` simply generates an external view of the transformed data. The actual purpose of the `STACK` procedures is to assemble a list of transformations for embedding in a model. The transformations are passed to `CREATE_MODEL` in the `xform_list` parameter. See INSERT_BIN_NUM_EQWIDTH Procedure for an example.

```
describe mining_data
 Name                                                      Null?    Type
 --------------------------------------------------------- -------- -----
 CUST_ID                                                   NOT NULL NUMBER
 CUST_CREDIT_LIMIT                                                  NUMBER

BEGIN
   dbms_data_mining_transform.create_miss_num ('miss_num_tbl');
   dbms_data_mining_transform.insert_miss_num_mean ('miss_num_tbl','mining_data');
END;
/
SELECT * FROM miss_num_tbl;

COL                  ATT      VAL
-------------------- ----- ------
CUST_ID                        5.5
CUST_CREDIT_LIMIT          185.71

DECLARE
    MINING_DATA_STACK   dbms_data_mining_transform.TRANSFORM_LIST;
  BEGIN
    dbms_data_mining_transform.STACK_MISS_NUM (
          miss_table_name   => 'miss_num_tbl',
          xform_list        => mining_data_stack);
    dbms_data_mining_transform.XFORM_STACK (
          xform_list        =>  mining_data_stack,
          data_table_name   => 'mining_data',
          xform_view_name   => 'mining_data_stack_view');
END;
/
-- Before transformation
SELECT * FROM mining_data
  ORDER BY cust_id;
CUST_ID CUST_CREDIT_LIMIT
------- -----------------
      1               100
      2
      3               200
      4
      5               150
      6               400
      7               150
      8
      9               100
     10               200

-- After transformation
SELECT * FROM mining_data_stack_view
  ORDER BY cust_id;
```

```
CUST_ID CUST_CREDIT_LIMIT
------- -----------------
      1               100
      2            185.71
      3               200
      4            185.71
      5               150
      6               400
      7               150
      8            185.71
      9               100
     10               200
```

STACK_NORM_LIN Procedure

This procedure adds linear normalization transformations to a transformation list.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.STACK_NORM_LIN (
     norm_table_name     IN       VARCHAR2,
     xform_list          IN OUT   NOCOPY TRANSFORM_LIST,
     norm_schema_name    IN       VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-40 STACK_NORM_LIN Procedure Parameters

Parameter	Description
norm_table_name	Name of the transformation definition table for linear normalization. You can use the CREATE_NORM_LIN Procedure to create the definition table. The table must be populated with transformation definitions before you call STACK_NORM_LIN.To populate the table, you can use one of the INSERT procedures for normalization or you can write your own SQL. See Table 63-16.
xform_list	A transformation list. See Table 63-1 for a description of the TRANSFORM_LIST object type.
norm_schema_name	Schema of *norm_table_name*. If no schema is specified, the current schema is used.

Usage Notes

See "Operational Notes". The following sections are especially relevant:

- "About Transformation Lists"
- "About Stacking"
- "Nested Data Transformations"

Examples

This example shows how the column cust_credit_limit could be normalized in a transformation list called mining_data_stack.

> **Note:**
>
> This example invokes the XFORM_STACK Procedure to show how the data is transformed by the stack. `XFORM_STACK` simply generates an external view of the transformed data. The actual purpose of the `STACK` procedures is to assemble a list of transformations for embedding in a model. The transformations are passed to `CREATE_MODEL` in the `xform_list` parameter. See INSERT_BIN_NUM_EQWIDTH Procedure for an example.

```
CREATE OR REPLACE VIEW mining_data AS
       SELECT cust_id, country_id, cust_postal_code, cust_credit_limit
          FROM sh.customers;
BEGIN
   dbms_data_mining_transform.create_norm_lin ('norm_lin_tbl');
   dbms_data_mining_transform.insert_norm_lin_minmax (
       norm_table_name   => 'norm_lin_tbl',
       data_table_name   => 'mining_data',
       exclude_list      =>  dbms_data_mining_transform.COLUMN_LIST('cust_id',
                                                                    'country_id'));
END;
/
SELECT * FROM norm_lin_tbl;
COL                  ATT    SHIFT  SCALE
-------------------- ----- ------ ------
CUST_CREDIT_LIMIT            1500  13500

DECLARE
   MINING_DATA_STACK   dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
   dbms_data_mining_transform.stack_norm_lin (
        norm_table_name   => 'norm_lin_tbl',
        xform_list        => mining_data_stack);
   dbms_data_mining_transform.XFORM_STACK (
        xform_list        =>  mining_data_stack,
        data_table_name   => 'mining_data',
        xform_view_name   => 'mining_data_stack_view');
END;
/
SELECT * FROM mining_data
  WHERE cust_id between 1 and 10
  ORDER BY cust_id;
CUST_ID COUNTRY_ID CUST_POSTAL_CODE     CUST_CREDIT_LIMIT
------- ---------- -------------------- -----------------
      1      52789 30828                            9000
      2      52778 86319                           10000
      3      52770 88666                            1500
      4      52770 87551                            1500
      5      52789 59200                            1500
      6      52769 77287                            1500
      7      52790 38763                            1500
      8      52790 58488                            3000
      9      52770 63033                            3000
     10      52790 52602                            3000

SELECT * FROM mining_data_stack_view
  WHERE cust_id between 1 and 10
  ORDER BY cust_id;
CUST_ID COUNTRY_ID CUST_POSTAL_CODE     CUST_CREDIT_LIMIT
```

```
------- ---------- ------------------- -----------------
      1      52789 30828                           .55556
      2      52778 86319                           .62963
      3      52770 88666                                0
      4      52770 87551                                0
      5      52789 59200                                0
      6      52769 77287                                0
      7      52790 38763                                0
      8      52790 58488                           .11111
      9      52770 63033                           .11111
     10      52790 52602                           .11111
```

XFORM_BIN_CAT Procedure

This procedure creates a view that implements the categorical binning transformations specified in a definition table. Only the columns that are specified in the definition table are transformed; the remaining columns from the data table are present in the view, but they are not changed.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.XFORM_BIN_CAT (
     bin_table_name       IN VARCHAR2,
     data_table_name      IN VARCHAR2,
     xform_view_name      IN VARCHAR2,
     literal_flag         IN BOOLEAN DEFAULT FALSE,
     bin_schema_name      IN VARCHAR2 DEFAULT NULL,
     data_schema_name     IN VARCHAR2 DEFAULT NULL,
     xform_schema_name    IN VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-41 XFORM_BIN_CAT Procedure Parameters

Parameter	Description
bin_table_name	Name of the transformation definition table for categorical binning. You can use the CREATE_BIN_CAT Procedure to create the definition table. The table must be populated with transformation definitions before you call XFORM_BIN_CAT. To populate the table, you can use one of the INSERT procedures for categorical binning or you can write your own SQL. See Table 63-4.
data_table_name	Name of the table containing the data to be transformed.
xform_view_name	Name of the view to be created. The view presents columns in *data_table_name* with the transformations specified in *bin_table_name*.
literal_flag	Indicates whether the values in the bin column in the transformation definition table are valid SQL literals. When *literal_flag* is FALSE (the default), the bin identifiers will be transformed to SQL literals by surrounding them with single quotes. Set *literal_flag* to TRUE if the bin identifiers are numbers that should have a numeric datatype, as is the case for an O-Cluster model. See "INSERT_BIN_NUM_EQWIDTH Procedure" for an example.
bin_schema_name	Schema of *bin_table_name*. If no schema is specified, the current schema is used.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.

Table 63-41 (Cont.) XFORM_BIN_CAT Procedure Parameters

Parameter	Description
xform_schema_name	Schema of *xform_view_name*. If no schema is specified, the current schema is used.

Usage Notes

See "Operational Notes".

Examples

This example creates a view that bins the `cust_postal_code` column. The data source consists of three columns from `sh.customer`.

```
describe mining_data
 Name                                      Null?    Type
 ----------------------------------------- -------- ------------------------
 CUST_ID                                   NOT NULL NUMBER
 CUST_POSTAL_CODE                          NOT NULL VARCHAR2(10)
 CUST_CREDIT_LIMIT                                  NUMBER

SELECT * FROM mining_data WHERE cust_id between 104066 and 104069;

   CUST_ID CUST_POSTAL_CODE
CUST_CREDIT_LIMIT
--------- --------------------
-----------------
    104066 69776
7000
    104067 52602
9000
    104068 55787
11000
    104069 55977
5000

BEGIN
  dbms_data_mining_transform.create_bin_cat(
     bin_table_name     => 'bin_cat_tbl');
  dbms_data_mining_transform.insert_bin_cat_freq(
     bin_table_name     => 'bin_cat_tbl',
     data_table_name    => 'mining_data',
     bin_num          => 10);
   dbms_data_mining_transform.xform_bin_cat(
     bin_table_name     => 'bin_cat_tbl',
     data_table_name    => 'mining_data',
     xform_view_name    => 'bin_cat_view');
END;
/

SELECT * FROM bin_cat_view WHERE cust_id between 104066 and 104069;

   CUST_ID CUST_POSTAL_CODE
CUST_CREDIT_LIMIT
---------- --------------------
-----------------
    104066 6
7000
```

```
    104067 11
9000
    104068 3
11000
    104069 11
5000

SELECT text FROM user_views WHERE view_name IN 'BIN_CAT_VIEW';

TEXT

--------------------------------------------------------------------------------

SELECT
"CUST_ID",DECODE("CUST_POSTAL_CODE",'38082','1','45704','9','48346','5','

55787','3','63736','2','67843','7','69776','6','72860','10','78558','4','80841',

'8',NULL,NULL,'11') "CUST_POSTAL_CODE","CUST_CREDIT_LIMIT" FROM
mining_data
```

XFORM_BIN_NUM Procedure

This procedure creates a view that implements the numerical binning transformations specified in a definition table. Only the columns that are specified in the definition table are transformed; the remaining columns from the data table are present in the view, but they are not changed.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.XFORM_BIN_NUM (
     bin_table_name     IN VARCHAR2,
     data_table_name    IN VARCHAR2,
     xform_view_name    IN VARCHAR2,
     literal_flag       IN BOOLEAN DEFAULT FALSE,
     bin_schema_name    IN VARCHAR2 DEFAULT NULL,
     data_schema_name   IN VARCHAR2 DEFAULT NULL,
     xform_schema_name  IN VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-42 XFORM_BIN_NUM Procedure Parameters

Parameter	Description
bin_table_name	Name of the transformation definition table for numerical binning. You can use the CREATE_BIN_NUM Procedure to create the definition table. The table must be populated with transformation definitions before you call XFORM_BIN_NUM. To populate the table, you can use one of the INSERT procedures for numerical binning or you can write your own SQL. See "Table 63-6".
data_table_name	Name of the table containing the data to be transformed
xform_view_name	Name of the view to be created. The view presents columns in *data_table_name* with the transformations specified in *bin_table_name*.

Table 63-42 (Cont.) XFORM_BIN_NUM Procedure Parameters

Parameter	Description
literal_flag	Indicates whether the values in the bin column in the transformation definition table are valid SQL literals. When *literal_flag* is FALSE (the default), the bin identifiers will be transformed to SQL literals by surrounding them with single quotes. Set *literal_flag* to TRUE if the bin identifiers are numbers that should have a numeric datatype, as is the case for an O-Cluster model. See "INSERT_BIN_NUM_EQWIDTH Procedure" for an example.
bin_schema_name	Schema of *bin_table_name*. If no schema is specified, the current schema is used.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.
xform_schema_name	Schema of *xform_view_name*. If no schema is specified, the current schema is used.

Usage Notes

See "Operational Notes".

Examples

This example creates a view that bins the cust_credit_limit column. The data source consists of three columns from sh.customer.

```
describe mining_data
 Name                                      Null?    Type
 ----------------------------------------- -------- ----------------------
 CUST_ID                                   NOT NULL NUMBER
 CUST_POSTAL_CODE                          NOT NULL VARCHAR2(10)
 CUST_CREDIT_LIMIT                                  NUMBER

column cust_credit_limit off
SELECT * FROM mining_data WHERE cust_id between 104066 and 104069;

   CUST_ID CUST_POSTAL_CODE   CUST_CREDIT_LIMIT
---------  ------------------
--------------------
    104066 69776                          7000
    104067 52602                          9000
    104068 55787                         11000
    104069 55977                          5000

BEGIN
   dbms_data_mining_transform.create_bin_num(
           bin_table_name     => 'bin_num_tbl');
   dbms_data_mining_transform.insert_autobin_num_eqwidth(
           bin_table_name     => 'bin_num_tbl',
           data_table_name    => 'mining_data',
           bin_num                => 5,
           max_bin_num            => 10,
           exclude_list       => dbms_data_mining_transform.COLUMN_LIST('cust_id'));
  dbms_data_mining_transform.xform_bin_num(
          bin_table_name      => 'bin_num_tbl',
          data_table_name     => 'mining_data',
          xform_view_name     => 'mining_data_view');
```

```
END;
/
describe mining_data_view
 Name                                      Null?    Type
 ----------------------------------------- -------- ----------------------
 CUST_ID                                   NOT NULL NUMBER
 CUST_POSTAL_CODE                          NOT NULL VARCHAR2(10)
 CUST_CREDIT_LIMIT                                  VARCHAR2(2)

col cust_credit_limit on
col cust_credit_limit format a25
SELECT * FROM mining_data_view WHERE cust_id between 104066 and 104069;

   CUST_ID CUST_POSTAL_CODE
CUST_CREDIT_LIMIT
---------- --------------------
-------------------------
    104066 69776
5
    104067 52602
6
    104068 55787
8
    104069 55977
3

set long 2000
SELECT text FROM user_views WHERE view_name IN 'MINING_DATA_VIEW';

TEXT

--------------------------------------------------------------------------------

SELECT "CUST_ID","CUST_POSTAL_CODE",CASE WHEN "CUST_CREDIT_LIMIT"<1500 THEN
NULL
 WHEN "CUST_CREDIT_LIMIT"<=2850 THEN '1' WHEN "CUST_CREDIT_LIMIT"<=4200 THEN
'2'
 WHEN "CUST_CREDIT_LIMIT"<=5550 THEN '3' WHEN "CUST_CREDIT_LIMIT"<=6900 THEN
'4'
 WHEN "CUST_CREDIT_LIMIT"<=8250 THEN '5' WHEN "CUST_CREDIT_LIMIT"<=9600 THEN
'6'
 WHEN "CUST_CREDIT_LIMIT"<=10950 THEN '7' WHEN "CUST_CREDIT_LIMIT"<=12300 THEN
'
8' WHEN "CUST_CREDIT_LIMIT"<=13650 THEN '9' WHEN "CUST_CREDIT_LIMIT"<=15000
THEN
 '10' END "CUST_CREDIT_LIMIT" FROM mining_data
```

XFORM_CLIP Procedure

This procedure creates a view that implements the clipping transformations specified in a definition table. Only the columns that are specified in the definition table are transformed; the remaining columns from the data table are present in the view, but they are not changed.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.XFORM_CLIP (
    clip_table_name       IN VARCHAR2,
    data_table_name       IN VARCHAR2,
    xform_view_name       IN VARCHAR2,
    clip_schema_name      IN VARCHAR2 DEFAULT NULL,
```

```
    data_schema_name      IN VARCHAR2,DEFAULT NULL,
    xform_schema_name     IN VARCHAR2,DEFAULT NULL);
```

Parameters

Table 63-43 XFORM_CLIP Procedure Parameters

Parameter	Description
clip_table_name	Name of the transformation definition table for clipping. You can use the CREATE_CLIP Procedure to create the definition table. The table must be populated with transformation definitions before you call XFORM_CLIP. To populate the table, you can use one of the INSERT procedures for clipping you can write your own SQL. See Table 63-8.
data_table_name	Name of the table containing the data to be transformed
xform_view_name	Name of the view to be created. The view presents columns in *data_table_name* with the transformations specified in *clip_table_name*.
clip_schema_name	Schema of *clip_table_name*. If no schema is specified, the current schema is used.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.
xform_schema_name	Schema of *xform_view_name*. If no schema is specified, the current schema is used.

Examples

This example creates a view that clips the cust_credit_limit column. The data source consists of three columns from sh.customer.

```
describe mining_data
 Name                           Null?    Type
 ------------------------------ -------- ------------------------
 CUST_ID                        NOT NULL NUMBER
 CUST_POSTAL_CODE               NOT NULL VARCHAR2(10)
 CUST_CREDIT_LIMIT                       NUMBER

BEGIN
   dbms_data_mining_transform.create_clip(
      clip_table_name    => 'clip_tbl');
   dbms_data_mining_transform.insert_clip_trim_tail(
      clip_table_name   => 'clip_tbl',
      data_table_name   => 'mining_data',
      tail_frac         => 0.05,
      exclude_list      => dbms_data_mining_transform.COLUMN_LIST('cust_id'));
   dbms_data_mining_transform.xform_clip(
      clip_table_name     => 'clip_tbl',
      data_table_name    => 'mining_data',
      xform_view_name    => 'clip_view');
END;
/
describe clip_view
 Name                           Null?    Type
 ------------------------------ -------- ------------------------
 CUST_ID                        NOT NULL NUMBER
 CUST_POSTAL_CODE               NOT NULL VARCHAR2(10)
 CUST_CREDIT_LIMIT                       NUMBER
```

```
SELECT MIN(cust_credit_limit), MAX(cust_credit_limit) FROM mining_data;

MIN(CUST_CREDIT_LIMIT) MAX(CUST_CREDIT_LIMIT)
---------------------- ----------------------
                  1500                  15000

SELECT MIN(cust_credit_limit), MAX(cust_credit_limit) FROM clip_view;

MIN(CUST_CREDIT_LIMIT) MAX(CUST_CREDIT_LIMIT)
---------------------- ----------------------
                  1500                  11000

set long 2000
SELECT text FROM user_views WHERE view_name IN 'CLIP_VIEW';

TEXT
--------------------------------------------------------------------------------
SELECT "CUST_ID","CUST_POSTAL_CODE",CASE WHEN "CUST_CREDIT_LIMIT" < 1500 THEN NU
LL WHEN "CUST_CREDIT_LIMIT" > 11000 THEN NULL ELSE "CUST_CREDIT_LIMIT" END "CUST
_CREDIT_LIMIT" FROM mining_data
```

XFORM_COL_REM Procedure

This procedure creates a view that implements the column removal transformations specified in a definition table. Only the columns that are specified in the definition table are removed; the remaining columns from the data table are present in the view.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.XFORM_COL_REM (
     rem_table_name     IN       VARCHAR2,
     data_table_name    IN       VARCHAR2,
     xform_view_name    IN       VARCHAR2,
     rem_schema_name    IN       VARCHAR2 DEFAULT NULL,
     data_schema_name   IN       VARCHAR2 DEFAULT NULL,
     xform_schema_name  IN       VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-44 XFORM_COL_REM Procedure Parameters

Parameter	Description
rem_table_name	Name of the transformation definition table for column removal. You can use the CREATE_COL_REM Procedure to create the definition table. See Table 63-10. The table must be populated with column names before you call XFORM_COL_REM. The INSERT_BIN_SUPER Procedure and the INSERT_AUTOBIN_NUM_EQWIDTH Procedure can optionally be used to populate the table. You can also use SQL INSERT statements.
data_table_name	Name of the table containing the data to be transformed
xform_view_name	Name of the view to be created. The view presents the columns in *data_table_name* that are not specified in *rem_table_name*.
rem_schema_name	Schema of *rem_table_name*. If no schema is specified, the current schema is used.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.

Table 63-44 (Cont.) XFORM_COL_REM Procedure Parameters

Parameter	Description
xform_schema_name	Schema of *xform_view_name*. If no schema is specified, the current schema is used.

Usage Notes

See "Operational Notes".

Examples

This example creates a view that includes all but one column from the table `customers` in the current schema.

```
describe customers
 Name                                    Null?    Type
 ----------------------------------------- -------- ----------------------------
 CUST_ID                                 NOT NULL NUMBER
 CUST_MARITAL_STATUS                              VARCHAR2(20)
 OCCUPATION                                       VARCHAR2(21)
 AGE                                              NUMBER
 YRS_RESIDENCE                                    NUMBER

BEGIN
    DBMS_DATA_MINING_TRANSFORM.CREATE_COL_REM ('colrem_xtbl');
END;
 /
INSERT INTO colrem_xtbl VALUES('CUST_MARITAL_STATUS', null);

NOTE: This currently doesn't work. See bug 9310319

BEGIN
   DBMS_DATA_MINING_TRANSFORM.XFORM_COL_REM (
     rem_table_name        => 'colrem_xtbl',
     data_table_name       => 'customers',
     xform_view_name       => 'colrem_view');
END;
/
describe colrem_view

 Name                                    Null?    Type
 ----------------------------------------- -------- ----------------------------
 CUST_ID                                 NOT NULL NUMBER
 OCCUPATION                                       VARCHAR2(21)
 AGE                                              NUMBER
 YRS_RESIDENCE                                    NUMBER
```

XFORM_EXPR_NUM Procedure

This procedure creates a view that implements the specified numeric transformations. Only the columns that you specify are transformed; the remaining columns from the data table are present in the view, but they are not changed.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.XFORM_EXPR_NUM (
     expr_pattern       IN       VARCHAR2,
```

```
data_table_name    IN       VARCHAR2,
xform_view_name    IN       VARCHAR2,
exclude_list       IN       COLUMN_LIST DEFAULT NULL,
include_list       IN       COLUMN_LIST DEFAULT NULL,
col_pattern        IN       VARCHAR2 DEFAULT ':col',
data_schema_name   IN       VARCHAR2 DEFAULT NULL,
xform_schema_name  IN       VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-45 XFORM_EXPR_NUM Procedure Parameters

Parameter	Description
expr_pattern	A numeric transformation expression
data_table_name	Name of the table containing the data to be transformed
xform_view_name	Name of the view to be created. The view presents columns in *data_table_name* with the transformations specified in *expr_pattern* and *col_pattern*.
exclude_list	List of numerical columns to exclude. If NULL, no numerical columns are excluded. The format of *exclude_list* is: `dbms_data_mining_transform.COLUMN_LIST('col1','col2', ...'coln')`
include_list	List of numeric columns to include. If NULL, all numeric columns are included. The format of *include_list* is: `dbms_data_mining_transform.COLUMN_LIST('col1','col2', ...'coln')`
col_pattern	The value within *expr_pattern* that will be replaced with a column name. The value of *col_pattern* is case-sensitive. The default value of *col_pattern* is ':col'
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.
xform_schema_name	Schema of *xform_view_name*. If no schema is specified, the current schema is used.

Usage Notes

1. The XFORM_EXPR_NUM procedure constructs numeric transformation expressions from the specified expression pattern (*expr_pattern*) by replacing every occurrence of the specified column pattern (*col_pattern*) with an actual column name.

   XFORM_EXPR_NUM uses the SQL REPLACE function to construct the transformation expressions.

   ```
   REPLACE (expr_pattern,col_pattern,'"column_name"') || '"column_name"'
   ```

   If there is a column match, then the replacement is made in the transformation expression; if there is not a match, then the column is used without transformation.

> **See:**
>
> *Oracle Database SQL Language Reference* for information about the `REPLACE` function

2. Because of the include and exclude list parameters, the `XFORM_EXPR_NUM` and `XFORM_EXPR_STR` procedures allow you to easily specify individual columns for transformation within large data sets. The other `XFORM_*` procedures support an exclude list only. In these procedures, you must enumerate every column that you do not want to transform.
3. See "Operational Notes"

Examples

This example creates a view that transforms the datatype of numeric columns.

```
describe customers
 Name                              Null?    Type
 --------------------------------- -------- ------------------------
 CUST_ID                           NOT NULL NUMBER
 CUST_MARITAL_STATUS                        VARCHAR2(20)
 OCCUPATION                                 VARCHAR2(21)
 AGE                                        NUMBER
 YRS_RESIDENCE                              NUMBER

BEGIN
  DBMS_DATA_MINING_TRANSFORM.XFORM_EXPR_NUM(
    expr_pattern         => 'to_char(:col)',
    data_table_name      => 'customers',
    xform_view_name      => 'cust_nonum_view',
    exclude_list         => dbms_data_mining_transform.COLUMN_LIST( 'cust_id'),
    include_list         => null,
    col_pattern          => ':col');
END;
/
describe cust_nonum_view
 Name                              Null?    Type
 --------------------------------- -------- ------------------------
 CUST_ID                           NOT NULL NUMBER
 CUST_MARITAL_STATUS                        VARCHAR2(20)
 OCCUPATION                                 VARCHAR2(21)
 AGE                                        VARCHAR2(40)
 YRS_RESIDENCE                              VARCHAR2(40)
```

XFORM_EXPR_STR Procedure

This procedure creates a view that implements the specified categorical transformations. Only the columns that you specify are transformed; the remaining columns from the data table are present in the view, but they are not changed.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.XFORM_EXPR_STR (
     expr_pattern       IN       VARCHAR2,
     data_table_name    IN       VARCHAR2,
     xform_view_name    IN       VARCHAR2,
     exclude_list       IN       COLUMN_LIST DEFAULT NULL,
     include_list       IN       COLUMN_LIST DEFAULT NULL,
```

```
    col_pattern        IN       VARCHAR2 DEFAULT ':col',
    data_schema_name   IN       VARCHAR2 DEFAULT NULL,
    xform_schema_name  IN       VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-46 XFORM_EXPR_STR Procedure Parameters

Parameter	Description
expr_pattern	A character transformation expression
data_table_name	Name of the table containing the data to be transformed
xform_view_name	Name of the view to be created. The view presents columns in *data_table_name* with the transformations specified in *expr_pattern* and *col_pattern*.
exclude_list	List of categorical columns to exclude. If NULL, no categorical columns are excluded. The format of *exclude_list* is: `dbms_data_mining_transform.COLUMN_LIST('col1','col2', ...'coln')`
include_list	List of character columns to include. If NULL, all character columns are included. The format of *include_list* is: `dbms_data_mining_transform.COLUMN_LIST('col1','col2', ...'coln')`
col_pattern	The value within *expr_pattern* that will be replaced with a column name. The value of *col_pattern* is case-sensitive. The default value of *col_pattern* is ':col'
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.
xform_schema_name	Schema of *xform_view_name*. If no schema is specified, the current schema is used.

Usage Notes

1. The XFORM_EXPR_STR procedure constructs character transformation expressions from the specified expression pattern (*expr_pattern*) by replacing every occurrence of the specified column pattern (*col_pattern*) with an actual column name.

   XFORM_EXPR_STR uses the SQL REPLACE function to construct the transformation expressions.

   ```
   REPLACE (expr_pattern,col_pattern,'"column_name"') || '"column_name"'
   ```

   If there is a column match, then the replacement is made in the transformation expression; if there is not a match, then the column is used without transformation.

> **See:**
>
> *Oracle Database SQL Language Reference* for information about the `REPLACE` function

2. Because of the include and exclude list parameters, the `XFORM_EXPR_STR` and `XFORM_EXPR_NUM` procedures allow you to easily specify individual columns for transformation within large data sets. The other `XFORM_*` procedures support an exclude list only. In these procedures, you must enumerate every column that you do not want to transform.
3. See "Operational Notes"

Examples

This example creates a view that transforms character columns to upper case.

```
describe customers
 Name                                      Null?    Type
 ----------------------------------------- -------- ----------------------------
 CUST_ID                                   NOT NULL NUMBER
 CUST_MARITAL_STATUS                                VARCHAR2(20)
 OCCUPATION                                         VARCHAR2(21)
 AGE                                                NUMBER
 YRS_RESIDENCE                                      NUMBER

SELECT cust_id,  cust_marital_status, occupation FROM customers
    WHERE   cust_id > 102995
    ORDER BY cust_id desc;

CUST_ID CUST_MARITAL_STATUS  OCCUPATION
------- -------------------- --------------------
 103000 Divorc.              Cleric.
 102999 Married              Cleric.
 102998 Married              Exec.
 102997 Married              Exec.
 102996 NeverM               Other

BEGIN
  DBMS_DATA_MINING_TRANSFORM.XFORM_EXPR_STR(
       expr_pattern            => 'upper(:col)',
       data_table_name         => 'customers',
       xform_view_name         => 'cust_upcase_view');
END;
/
describe cust_upcase_view
 Name                           Null?    Type
 ------------------------------ -------- --------------------
 CUST_ID                        NOT NULL NUMBER
 CUST_MARITAL_STATUS                     VARCHAR2(20)
 OCCUPATION                              VARCHAR2(21)
 AGE                                     NUMBER
 YRS_RESIDENCE                           NUMBER

SELECT cust_id,  cust_marital_status, occupation FROM cust_upcase_view
   WHERE   cust_id > 102995
   ORDER BY cust_id desc;

CUST_ID CUST_MARITAL_STATUS  OCCUPATION
```

```
------- -------------------- --------------------
 103000 DIVORC.              CLERIC.
 102999 MARRIED              CLERIC.
 102998 MARRIED              EXEC.
 102997 MARRIED              EXEC.
 102996 NEVERM               OTHER
```

XFORM_MISS_CAT Procedure

This procedure creates a view that implements the categorical missing value treatment transformations specified in a definition table. Only the columns that are specified in the definition table are transformed; the remaining columns from the data table are present in the view, but they are not changed.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.XFORM_MISS_CAT (
    miss_table_name       IN VARCHAR2,
    data_table_name       IN VARCHAR2,
    xform_view_name       IN VARCHAR2,
    miss_schema_name      IN VARCHAR2 DEFAULT NULL,
    data_schema_name      IN VARCHAR2 DEFAULT NULL,
    xform_schema_name     IN VARCHAR2 DEFAULT NULL;
```

Parameters

Table 63-47 XFORM_MISS_CAT Procedure Parameters

Parameter	Description
miss_table_name	Name of the transformation definition table for categorical missing value treatment. You can use the CREATE_MISS_CAT Procedure to create the definition table. The table must be populated with transformation definitions before you call XFORM_MISS_CAT. To populate the table, you can use the INSERT_MISS_CAT_MODE Procedure or you can write your own SQL. See Table 63-12.
data_table_name	Name of the table containing the data to be transformed
xform_view_name	Name of the view to be created. The view presents columns in *data_table_name* with the transformations specified in *miss_table_name*.
miss_schema_name	Schema of *miss_table_name*. If no schema is specified, the current schema is used.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.
xform_schema_name	Schema of *xform_view_name*. If no schema is specified, the current schema is used.

Usage Notes

See "Operational Notes".

Examples

This example creates a view that replaces missing categorical values with the mode.

```
SELECT * FROM geog;

REG_ID REGION
```

```
------ ------------------------------
     1 NE
     2 SW
     3 SE
     4 SW
     5
     6 NE
     7 NW
     8 NW
     9
    10
    11 SE
    12 SE
    13 NW
    14 SE
    15 SE

SELECT STATS_MODE(region) FROM geog;

STATS_MODE(REGION)
------------------------------
SE

BEGIN
  DBMS_DATA_MINING_TRANSFORM.CREATE_MISS_CAT('misscat_xtbl');
  DBMS_DATA_MINING_TRANSFORM.INSERT_MISS_CAT_MODE (
    miss_table_name         => 'misscat_xtbl',
    data_table_name         => 'geog' );
END;
/

SELECT col, val FROM misscat_xtbl;

COL        VAL
---------- ----------
REGION     SE

BEGIN
  DBMS_DATA_MINING_TRANSFORM.XFORM_MISS_CAT (
     miss_table_name        => 'misscat_xtbl',
     data_table_name        => 'geog',
     xform_view_name        => 'geogxf_view');
END;
/

SELECT * FROM geogxf_view;

REG_ID REGION
------ ------------------------------
     1 NE
     2 SW
     3 SE
     4 SW
     5 SE
     6 NE
     7 NW
     8 NW
     9 SE
    10 SE
    11 SE
    12 SE
    13 NW
```

```
    14 SE
    15 SE
```

XFORM_MISS_NUM Procedure

This procedure creates a view that implements the numerical missing value treatment transformations specified in a definition table. Only the columns that are specified in the definition table are transformed; the remaining columns from the data table are present in the view, but they are not changed.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.XFORM_MISS_NUM (
    miss_table_name       IN VARCHAR2,
    data_table_name       IN VARCHAR2,
    xform_view_name       IN VARCHAR2,
    miss_schema_name      IN VARCHAR2 DEFAULT NULL,
    data_schema_name      IN VARCHAR2 DEFAULT NULL,
    xform_schema_name     IN VARCHAR2 DEFAULT NULL;
```

Parameters

Table 63-48 XFORM_MISS_NUM Procedure Parameters

Parameter	Description
miss_table_name	Name of the transformation definition table for numerical missing value treatment. You can use the CREATE_MISS_NUM Procedure to create the definition table. The table must be populated with transformation definitions before you call XFORM_MISS_NUM. To populate the table, you can use the INSERT_MISS_NUM_MEAN Procedure or you can write your own SQL. See Table 63-14.
data_table_name	Name of the table containing the data to be transformed
xform_view_name	Name of the view to be created. The view presents columns in *data_table_name* with the transformations specified in *miss_table_name*.
miss_schema_name	Schema of *miss_table_name*. If no schema is specified, the current schema is used.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.
xform_schema_name	Schema of *xform_view_name*. If no schema is specified, the current schema is used.

Usage Notes

See "Operational Notes".

Examples

This example creates a view that replaces missing numerical values with the mean.

```
SELECT * FROM items;

ITEM_ID       QTY
---------- ------
aa            200
```

```
bb            200
cc            250
dd
ee
ff            100
gg            250
hh            200
ii
jj            200

SELECT AVG(qty) FROM items;

AVG(QTY)
--------
     200

BEGIN
  DBMS_DATA_MINING_TRANSFORM.CREATE_MISS_NUM('missnum_xtbl');
  DBMS_DATA_MINING_TRANSFORM.INSERT_MISS_NUM_MEAN (
     miss_table_name          => 'missnum_xtbl',
     data_table_name         => 'items' );
END;
/

SELECT col, val FROM missnum_xtbl;

COL           VAL
---------- ------
QTY           200

BEGIN
    DBMS_DATA_MINING_TRANSFORM.XFORM_MISS_NUM (
        miss_table_name        => 'missnum_xtbl',
        data_table_name        => 'items',
        xform_view_name        => 'items_view');
END;
/

SELECT * FROM items_view;

ITEM_ID       QTY
---------- ------
aa            200
bb            200
cc            250
dd            200
ee            200
ff            100
gg            250
hh            200
ii            200
jj            200
```

XFORM_NORM_LIN Procedure

This procedure creates a view that implements the linear normalization transformations specified in a definition table. Only the columns that are specified in the definition table are

transformed; the remaining columns from the data table are present in the view, but they are not changed.

Syntax

```
DBMS_DATA_MINING_TRANSFORM.XFORM_NORM_LIN (
     norm_table_name      IN VARCHAR2,
     data_table_name      IN VARCHAR2,
     xform_view_name      IN VARCHAR2,
     norm_schema_name     IN VARCHAR2 DEFAULT NULL,
     data_schema_name     IN VARCHAR2 DEFAULT NULL,
     xform_schema_name    IN VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-49 XFORM_NORM_LIN Procedure Parameters

Parameter	Description
norm_table_name	Name of the transformation definition table for linear normalization. You can use the CREATE_NORM_LIN Procedure to create the definition table. The table must be populated with transformation definitions before you call XFORM_NORM_LIN. To populate the table, you can use one of the INSERT procedures for normalization or you can write your own SQL. See Table 63-12.
data_table_name	Name of the table containing the data to be transformed
xform_view_name	Name of the view to be created. The view presents columns in *data_table_name* with the transformations specified in *miss_table_name*.
norm_schema_name	Schema of *miss_table_name*. If no schema is specified, the current schema is used.
data_schema_name	Schema of *data_table_name*. If no schema is specified, the current schema is used.
xform_schema_name	Schema of *xform_view_name*. If no schema is specified, the current schema is used.

Usage Notes

See "Operational Notes".

Examples

This example creates a view that normalizes the cust_year_of_birth and cust_credit_limit columns. The data source consists of three columns from sh.customer.

```
CREATE OR REPLACE VIEW mining_data AS
     SELECT cust_id, cust_year_of_birth, cust_credit_limit
     FROM sh.customers;

describe mining_data
 Name                                      Null?    Type
 ----------------------------------------- -------- ----------------------------
 CUST_ID                                   NOT NULL NUMBER
 CUST_YEAR_OF_BIRTH                        NOT NULL NUMBER(4)
 CUST_CREDIT_LIMIT                                  NUMBER

SELECT * FROM mining_data WHERE cust_id > 104495
      ORDER BY cust_year_of_birth;
```

```
 CUST_ID CUST_YEAR_OF_BIRTH CUST_CREDIT_LIMIT
-------- ------------------ -----------------
  104496               1947              3000
  104498               1954             10000
  104500               1962             15000
  104499               1970              3000
  104497               1976              3000

BEGIN
  dbms_data_mining_transform.CREATE_NORM_LIN(
       norm_table_name        => 'normx_tbl');
 dbms_data_mining_transform.INSERT_NORM_LIN_MINMAX(
      norm_table_name     => 'normx_tbl',
      data_table_name     => 'mining_data',
      exclude_list     => dbms_data_mining_transform.COLUMN_LIST( 'cust_id'),
      round_num      => 3);
END;
/

SELECT col, shift, scale FROM normx_tbl;

COL                             SHIFT    SCALE
------------------------------ -------- --------
CUST_YEAR_OF_BIRTH                1910       77
CUST_CREDIT_LIMIT                 1500    13500

BEGIN
  DBMS_DATA_MINING_TRANSFORM.XFORM_NORM_LIN (
     norm_table_name     => 'normx_tbl',
     data_table_name     => 'mining_data',
     xform_view_name     => 'norm_view');
END;
/

SELECT * FROM norm_view WHERE cust_id > 104495
      ORDER BY cust_year_of_birth;

 CUST_ID CUST_YEAR_OF_BIRTH CUST_CREDIT_LIMIT
-------- ------------------ -----------------
  104496          .4805195          .1111111
  104498          .5714286          .6296296
  104500          .6753247                 1
  104499          .7792208          .1111111
  104497          .8571429          .1111111


set long 2000
SQL> SELECT text FROM user_views WHERE view_name IN 'NORM_VIEW';

TEXT
--------------------------------------------------------------------------------
SELECT "CUST_ID",("CUST_YEAR_OF_BIRTH"-1910)/77 "CUST_YEAR_OF_BIRTH",("CUST
_CREDIT_LIMIT"-1500)/13500 "CUST_CREDIT_LIMIT" FROM mining_data
```

XFORM_STACK Procedure

This procedure creates a view that implements the transformations specified by the stack. Only the columns and nested attributes that are specified in the stack are transformed. Any

remaining columns and nested attributes from the data table appear in the view without changes.

To create a list of objects that describe the transformed columns, use the DESCRIBE_STACK Procedure.

> **See Also:**
>
> "Overview"
>
> *Oracle Machine Learning for SQL User's Guide* for more information about machine learning attributes

Syntax

```
DBMS_DATA_MINING_TRANSFORM.XFORM_STACK (
     xform_list         IN     TRANSFORM_list,
     data_table_name    IN     VARCHAR2,
     xform_view_name    IN     VARCHAR2,
     data_schema_name   IN     VARCHAR2 DEFAULT NULL,
     xform_schema_name  IN     VARCHAR2 DEFAULT NULL);
```

Parameters

Table 63-50 XFORM_STACK Procedure Parameters

Parameter	Description
`xform_list`	The transformation list. See Table 63-1 for a description of the `TRANSFORM_LIST` object type.
`data_table_name`	Name of the table containing the data to be transformed
`xform_view_name`	Name of the view to be created. The view applies the transformations in *`xform_list`* to *`data_table_name`*.
`data_schema_name`	Schema of *`data_table_name`*. If no schema is specified, the current schema is used.
`xform_schema_name`	Schema of *`xform_view_name`*. If no schema is specified, the current schema is used.

Usage Notes

See "Operational Notes". The following sections are especially relevant:

- "About Transformation Lists"
- "About Stacking"
- "Nested Data Transformations"

Examples

This example applies a transformation list to the view `oml_user.cust_info` and shows how the data is transformed. The `CREATE` statement for `cust_info` is shown in "DESCRIBE_STACK Procedure".

```
BEGIN
  dbms_data_mining_transform.CREATE_BIN_NUM ('birth_yr_bins');
  dbms_data_mining_transform.INSERT_BIN_NUM_QTILE (
        bin_table_name   => 'birth_yr_bins',
        data_table_name  =>  'cust_info',
        bin_num          =>   6,
        exclude_list     =>  dbms_data_mining_transform.column_list(
                                 'cust_id','country_id'));
END;
/
SELECT * FROM birth_yr_bins;

COL                  ATT      VAL BIN
-------------------- ----- ------ ----------
CUST_YEAR_OF_BIRTH           1922
CUST_YEAR_OF_BIRTH           1951 1
CUST_YEAR_OF_BIRTH           1959 2
CUST_YEAR_OF_BIRTH           1966 3
CUST_YEAR_OF_BIRTH           1973 4
CUST_YEAR_OF_BIRTH           1979 5
CUST_YEAR_OF_BIRTH           1986 6

DECLARE
      cust_stack   dbms_data_mining_transform.TRANSFORM_LIST;
BEGIN
      dbms_data_mining_transform.SET_TRANSFORM (cust_stack,
          'country_id', NULL, 'country_id/10', 'country_id*10');
      dbms_data_mining_transform.STACK_BIN_NUM ('birth_yr_bins',
          cust_stack);
      dbms_data_mining_transform.SET_TRANSFORM (cust_stack,
          'custprods', 'Mouse Pad', 'value*100', 'value/100');
      dbms_data_mining_transform.XFORM_STACK(
          xform_list        => cust_stack,
          data_table_name   => 'cust_info',
          xform_view_name   =>  'cust_xform_view');
  END;
/

-- Two rows of data without transformations
SELECT * from cust_info WHERE cust_id BETWEEN 100010 AND 100011;

CUST_ID COUNTRY_ID CUST_YEAR_OF_BIRTH CUSTPRODS(ATTRIBUTE_NAME, VALUE)
------- ---------- ------------------ ----------------------------------------
 100010      52790               1975 DM_NESTED_NUMERICALS(
                                        DM_NESTED_NUMERICAL(
                                          '18" Flat Panel Graphics Monitor', 1),
                                        DM_NESTED_NUMERICAL(
                                          'SIMM- 16MB PCMCIAII card', 1))
 100011      52775               1972 DM_NESTED_NUMERICALS(
                                       DM_NESTED_NUMERICAL(
                                         'External 8X CD-ROM', 1),
                                       DM_NESTED_NUMERICAL(
                                         'Mouse Pad', 1),
                                       DM_NESTED_NUMERICAL(
                                        'SIMM- 16MB PCMCIAII card', 1),
                                       DM_NESTED_NUMERICAL(
                                         'Keyboard Wrist Rest', 1),
                                       DM_NESTED_NUMERICAL(
                                         '18" Flat Panel Graphics Monitor', 1),
                                       DM_NESTED_NUMERICAL(
                                         'O/S Documentation Set - English', 1))
```

```
-- Same two rows of data with transformations
SELECT * FROM cust_xform_view WHERE cust_id BETWEEN 100010 AND 100011;

CUST_ID  COUNTRY_ID  C  CUSTPRODS(ATTRIBUTE_NAME, VALUE)
-------  ----------  -  ------------------------------------------------------------
 100010        5279  5  DM_NESTED_NUMERICALS(
                           DM_NESTED_NUMERICAL(
                            '18" Flat Panel Graphics Monitor', 1),
                           DM_NESTED_NUMERICAL(
                             'SIMM- 16MB PCMCIAII card', 1))
 100011     5277.5  4  DM_NESTED_NUMERICALS(
                           DM_NESTED_NUMERICAL(
                             'External 8X CD-ROM', 1),
                           DM_NESTED_NUMERICAL(
                             'Mouse Pad', 100),
                           DM_NESTED_NUMERICAL(
                             'SIMM- 16MB PCMCIAII card', 1),
                           DM_NESTED_NUMERICAL(
                             'Keyboard Wrist Rest', 1),
                           DM_NESTED_NUMERICAL(
                             '18" Flat Panel Graphics Monitor', 1),
                           DM_NESTED_NUMERICAL(
                             'O/S Documentation Set - English', 1))
```