

The screenshot shows the Orange 3 interface with a workflow consisting of 'File', 'Preprocess', and 'Data Table' widgets. The 'Preprocess' widget is configured with the following settings:

- Preprocessors:**
 - Discretize Continuous Variables
 - Continue Discrete Variables
 - Impute Missing Values
 - Select Relevant Features
 - Select Random Features
 - Normalize Features
 - Randomize
 - Remove Sparse Features
 - Principal Component Analysis
- Impute Missing Values:**
 - ☒ Average/Most frequent
 - ☐ Replace with random value
 - ☐ Remove rows with missing values.
- Normalize Features:**
 - ☒ Standardize to $\mu=0$, $\sigma^2=1$
 - ☐ Center to $\mu=0$
 - ☐ Scale to $\sigma^2=1$
 - ☐ Normalize to interval $[0, 1]$

The 'Data Table (1)' widget displays a dataset with the following columns: Name, Doctor, Hospital, Age, and Gender. The data is as follows:

	Name	Doctor	Hospital	Age	Gender
1	Bobby JacksOn	Matthew Smith	Sons and Miller	30	Male
2	LesLie TErRy	Samantha Davies	Kim Inc	62	Male
3	DaNnY sMiTh	Tiffany Mitchell	Cook PLC	76	Female
4	andriw waTIS	Kevin Wells	Hernandez Rog...	28	Female
5	adriENNE bEll	Kathleen Hanna	White-White	43	Female
6	EMILY JOHNSOn	Taylor Newton	Nunez-Humphr...	36	Male
7	edwArD EDWa...	Kelly Olson	Group Middleton	21	Female
8	CHRISTinA MAR...	Suzanne Thomas	Powell Robinso...	20	Female
9	JASmiNe aGullaR	Daniel Ferguson	Sons Rich and	82	Male
10	CHRISToPher Be...	Heather Day	Padilla-Walker	58	Female
11	michElle danElS	John Duncan	Schaefer-Porter	72	Male
12	aaRon MARiNeZ	Douglas Mayo	Lyons-Blair	38	Female
13	connOR HANsEn	Kenneth Fletcher	Powers Miller, a...	75	Female
14	rObert bAuer	Theresa Freeman	Rivera-Gutierrez	68	Female
15	bROOKE brady	Roberta Stewart	Morris-Arellano	44	Female
16	MS. nAtAlie gA...	Maria Dougherty	Cline-Williams	46	Female
17	haley perkins	Erica Spencer	Cervantes-Wells	63	Female
18	mRS. jamiE cA...	Justin Kim	Torres, and Harr...	38	Male
19	LuKE BuRgEss	Justin Moore Jr.	Houston PLC	34	Female

2

The screenshot shows the 'Preprocess' widget settings in Orange 3. The 'Impute Missing Values' section is expanded, showing the following options:

- ☒ Average/Most frequent
- ☐ Replace with random value
- ☐ Remove rows with missing values.

The screenshot shows the 'Preprocess' widget settings in Orange 3. The 'Discretize Continuous Variables' section is expanded, showing the following options:

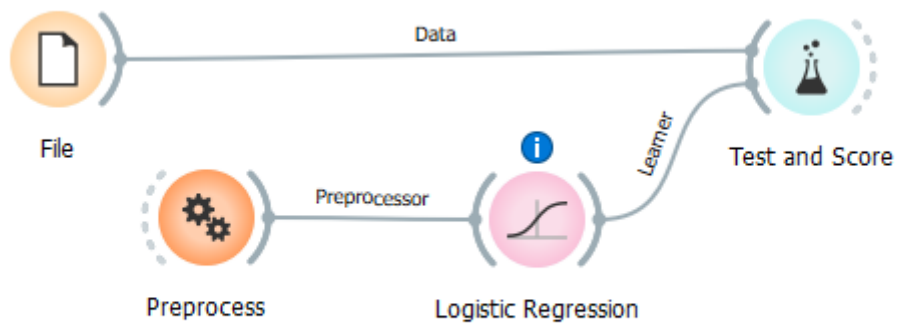
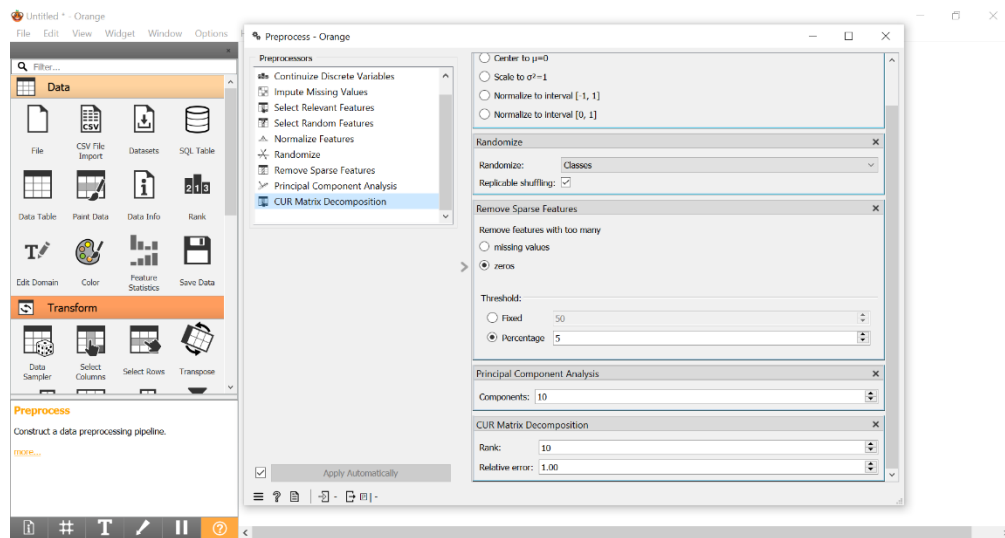
- ☐ Entropy-MDL discretization
- ☒ Equal frequency discretization
- ☐ Equal width discretization
- ☐ Remove numeric features

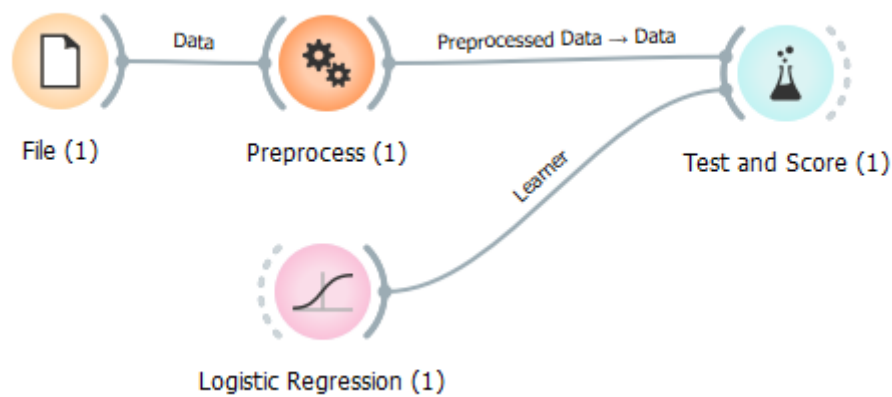
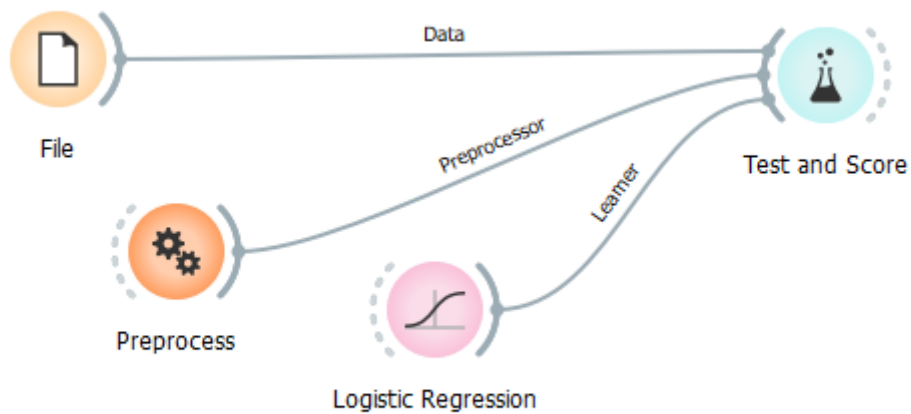
The 'Continue Discrete Variables' section is also expanded, showing the following options:

- ☐ Most frequent is base
- ☒ One feature per value
- ☐ Remove non binary features
- ☐ Remove categorical features
- ☐ Treat as ordinal
- ☐ Divide by number of values

The 'Select Relevant Features' section is also expanded, showing the following options:

- ☐ Score
- ☐ Information Gain
- ☐ Number of features
- ☒ Fixed: 10
- ☐ Proportion: 75.00%





Orange 3.10.0 - Untitled * - Orange

File Edit View Widget Window Options Help

Filter...

Data

File CSV File Import Datasets SQL Table

Data Table

Preprocess - Orange

Preprocessors

- Discretize Continuous Variables
- Continue Discrete Variables
- Impute Missing Values
- Select Relevant Features
- Normalize Features
- Remove Sparse Features
- Principal Component Analysis

Apply Automatically

Impute Missing Values

- ☒ Average/Most frequent
- ☐ Replace with random value
- ☐ Remove rows with missing values

Normalize Features

- ☒ Standardize to $\mu=0, \sigma^2=1$
- ☐ Center to $\mu=0$
- ☐ Scale to $\sigma^2=1$
- ☐ Normalize to interval [-1, 1]

Data Table - Orange

	Name	Doctor	Hospital	Age	Gender
1	Bobby JacksOn	Matthew Smith	Sons and Miller	30	Male
2	LesLie TErry	Samantha Davies	Kim Inc	62	Male
3	DaNnY sMitH	Tiffany Mitchell	Cook PLC	76	Female
4	andEw waITS	Kevin Wells	Hernandez Rog...	28	Female
5	adRIENNE bEIl	Kathleen Hanna	White-White	43	Female
6	EMILY JOHNSOn	Taylor Newton	Nunez-Humphr...	36	Male
7	edwArD EDWa...	Kelly Olson	Group Middleton	21	Female
8	CHrIStinA MAR...	Suzanne Thomas	Powell Robinso...	20	Female
9	JASmiNe aGullaR	Daniel Ferguson	Sons Rich and	82	Male
10	CHrIStopher Be...	Heather Day	Padilla-Walker	58	Female
11	mlchElLe daniELs	John Duncan	Schaefer-Porter	72	Male
12	aaRon MARTiNeZ	Douglas Mayo	Lyons-Blair	38	Female
13	connOR HANsEn	Kenneth Fletcher	Powers Miller, a...	75	Female
14	rObErT bAuer	Theresa Freeman	Rivera-Gutierrez	68	Female
15	bROOKE brady	Roberta Stewart	Morris-Arellano	44	Female
16	MS. nAtalie gA...	Maria Dougherty	Cline-Williams	46	Female
17	haley perkins	Erica Spencer	Cervantes-Wells	63	Female
18	mRS. jamie cA...	Justin Kim	Torres, and Harr...	38	Male
19	LuKE BuRgEss	Justin Moore Jr.	Houston PLC	34	Female

Untitled * - Orange

File Edit View Widget Window Options Help

Filter...

Data

File CSV File Import Datasets SQL Table

Data Table Paint Data Data Info Rank

Edit Domain Color Feature Statistics Save Data

Transform

Data Sampler Select Columns Select Rows Transpose

Test and Score

Cross validation accuracy estimation.

more...

Diagram: Datasets -> Preprocess -> Logistic Regression -> Test and Score

Test and Score - Orange

☒ Cross validation

Number of folds: 5

☒ Stratified

☐ Cross validation by feature

☐ Random sampling

Repeat train/test: 10

Training set size: 66 %

☒ Stratified

☐ Leave one out

Evaluation results for target

Model

Logistic Regression

Compare models by: Negligible diff.: 0.1

Logistic ...

Logistic Regression

Table shows probabilities that the score for the model in the row is higher than that of the model in the columns. Small numbers show the probability that the difference is negligible.

3

Untitled * - Orange

File Edit View Widget Window Options Help

Filter...

Data

File CSV File Import Datasets SQL Table

Data Table Paint Data Data Info Rank

Edit Domain Color Feature Statistics Save Data

Transform

Visualize

Model

Evaluate

k-Means

k-Means clustering algorithm with silhouette-based quality estimation.

more...

Diagram: k-Means

k-Means - Orange

Number of Clusters

☐ Fixed: 3

☒ From 2 to 8

Preprocessing

☒ Normalize columns

Initialization

Initialize with kMeans++

Re-runs: 10

Maximum iterations: 300

☒ Apply Automatically

Silhouette Scores

Number of Clusters	Silhouette Score
2	0.580
3	0.459
4	0.390
5	0.348
6	0.328
7	0.335
8	0.343

