

# Computer class

For Machine Learning

# Linear Regression

Target – Numeric variable

File

mult.tab

14 feature(s), 0 meta attribute(s)  
1 categorical class with 2 values.

Type	Role	Values
numeric	feature	
categorical	feature	Private, Self-emp...
numeric	feature	
categorical	feature	Bachelors, Some-...
numeric	feature	
categorical	feature	Married-civ-spou...

documentation data sets

Apply

Select Columns

Available Variables

Filter

- education
- workclass
- marital-status
- occupation
- relationship
- race
- sex
- native-country
- y

Features

- age
- fnlwgt
- education-num
- capital-gain
- capital-loss

Target Variable

- hours-per-week

Meta Attributes

Reset

Send Automatically

Data Table (1)

Info

6 instances (no missing values)  
1 feature (no missing values)  
No target variable.  
1 meta attribute (no missing values)

	name	value
1	intercept	31.900
2	age	0.0508
3	fnlwgt	-0.0000
4	education-n...	0.6486
5	capital-gain	0.0000
6	capital-loss	0.0012

Variables

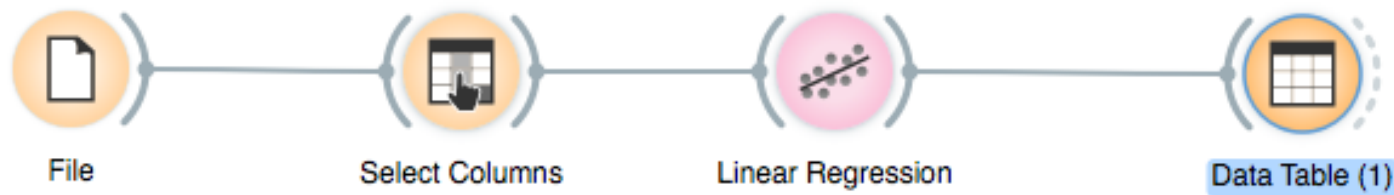
- Show variable labels (if present)
- Visualize numeric values
- Color by instance classes

Selection

- Select full rows

Restore Original Order

Send Automatically



Regression result in equation:

$$= \mathbf{31.9} + \mathbf{0.05} * x_1 + \mathbf{(-0.000001)} * x_2 + \mathbf{0.65} * x_3 + \mathbf{0.0001} * x_4 + \mathbf{0.001} *$$

Y – HOURS-PER-WEEK (weekly working hours)

$x_1$  – AGE (age of the respondent)

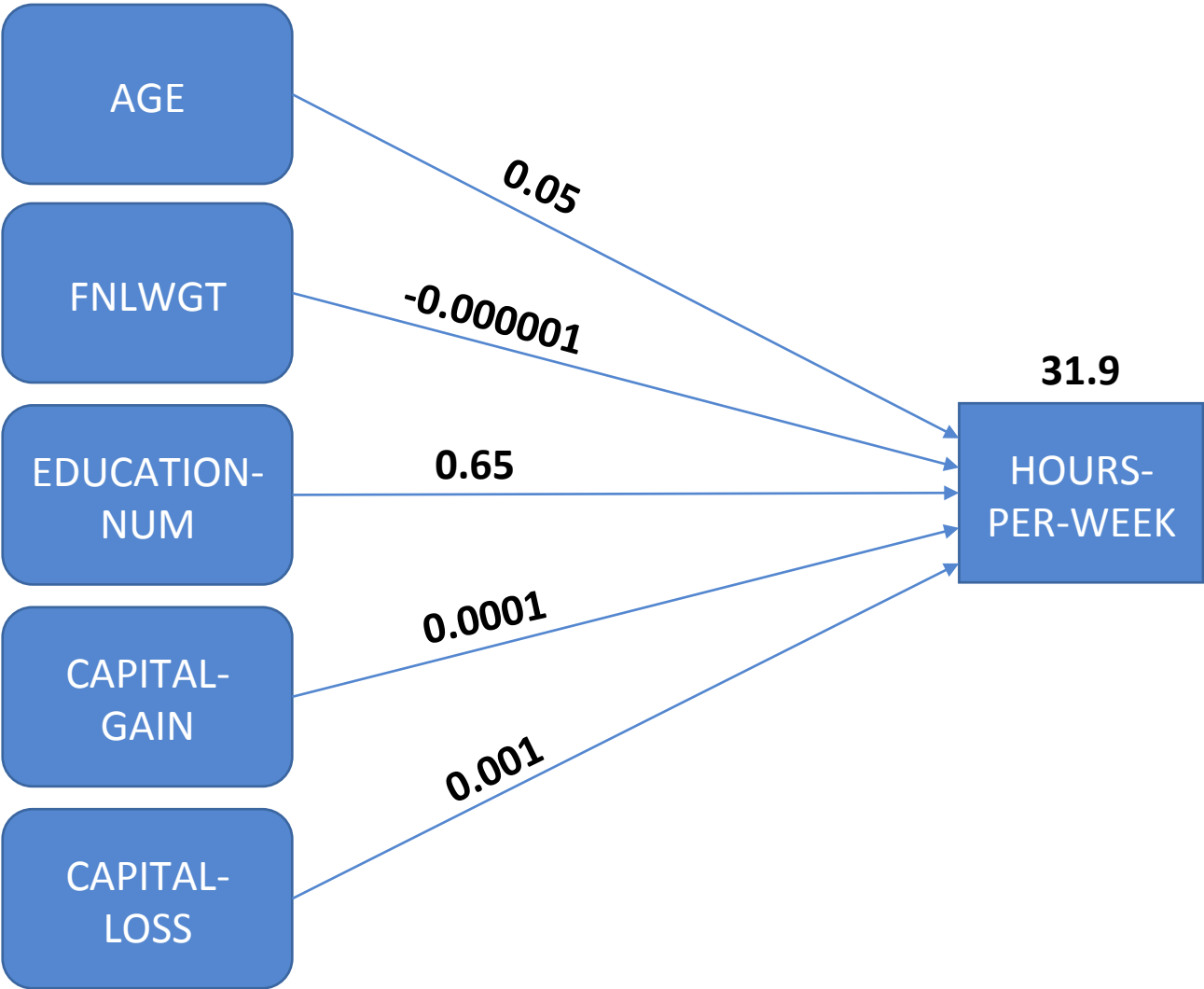
$x_2$  – FNLWGT (weight variable)

$x_3$  – EDUCATION-NUM (education degree)

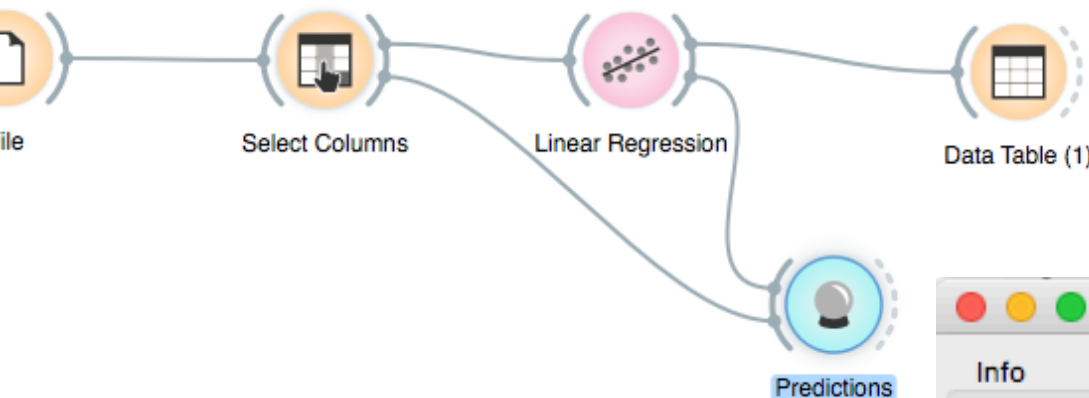
$x_4$  – CAPITAL-GAIN (income from investments)

$x_5$  – CAPITAL-LOSS (deficit from investments)

# Regression result in diagram:



# How good is the model?

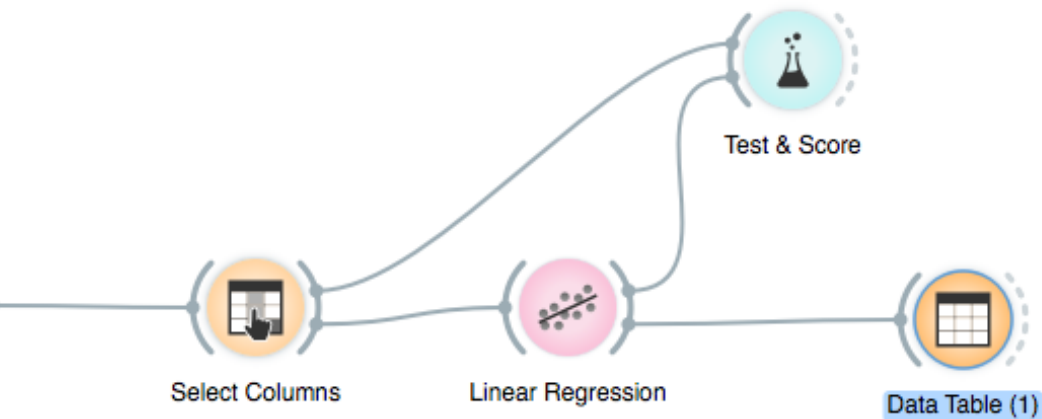


	Prediction	Real results	Predicti
Info	Linear Regression	hours-per-week	age
Data: 32561 instances.			
Predictors: 1			
Task: Regression			
<a href="#">Restore Original Order</a>			
Data View			
<input checked="" type="checkbox"/> Show full data set			
Output			
<input checked="" type="checkbox"/> Original data			
<input checked="" type="checkbox"/> Predictions			
<input checked="" type="checkbox"/> Probabilities			
1	42.456	40.000	39.000
2	42.796	13.000	50.000
3	39.464	40.000	38.000
4	38.911	40.000	53.000
5	41.432	40.000	28.000
6	42.590	40.000	37.000
7	37.482	16.000	49.000
8	40.182	45.000	52.000
9	43.899	50.000	31.000
10	42.825	40.000	42.000
11	39.999	80.000	37.000

# Testing Linear model (2)

RMSE > 0.5  
is bad fit

3% explain  
only



Test & Score

Sampling

- Cross validation
  - Number of folds: 10
  - Stratified
- Cross validation by feature
- Random sampling
  - Repeat train/test: 10
  - Training set size: 70 %
  - Stratified
- Leave one out
- Test on train data
- Test on test data

Evaluation Results

Method	MSE	RMSE	MAE	R2
Linear Regression	147.829	12.159	7.767	0.03

# Logistic Regression

Select Columns

Available Variables

Filter

- education
- workclass
- occupation
- relationship
- race
- native-country
- y
- fnlwgt

Features

- age
- education-num
- capital-gain
- capital-loss
- hours-per-week
- sex

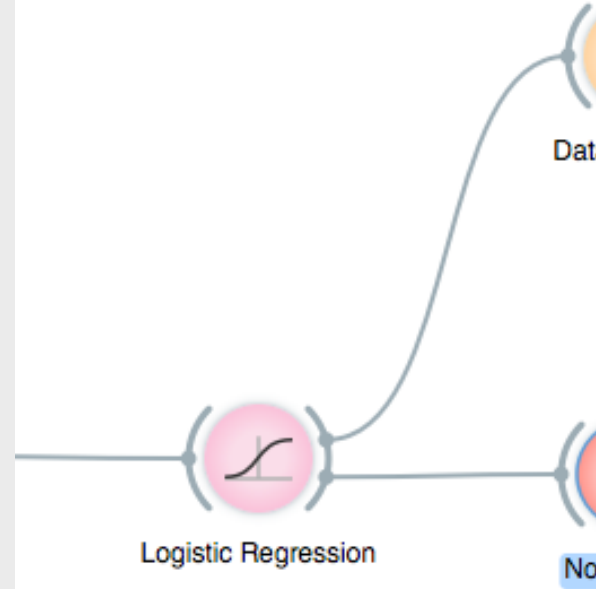
Target Variable

- marital-status

Meta Attributes

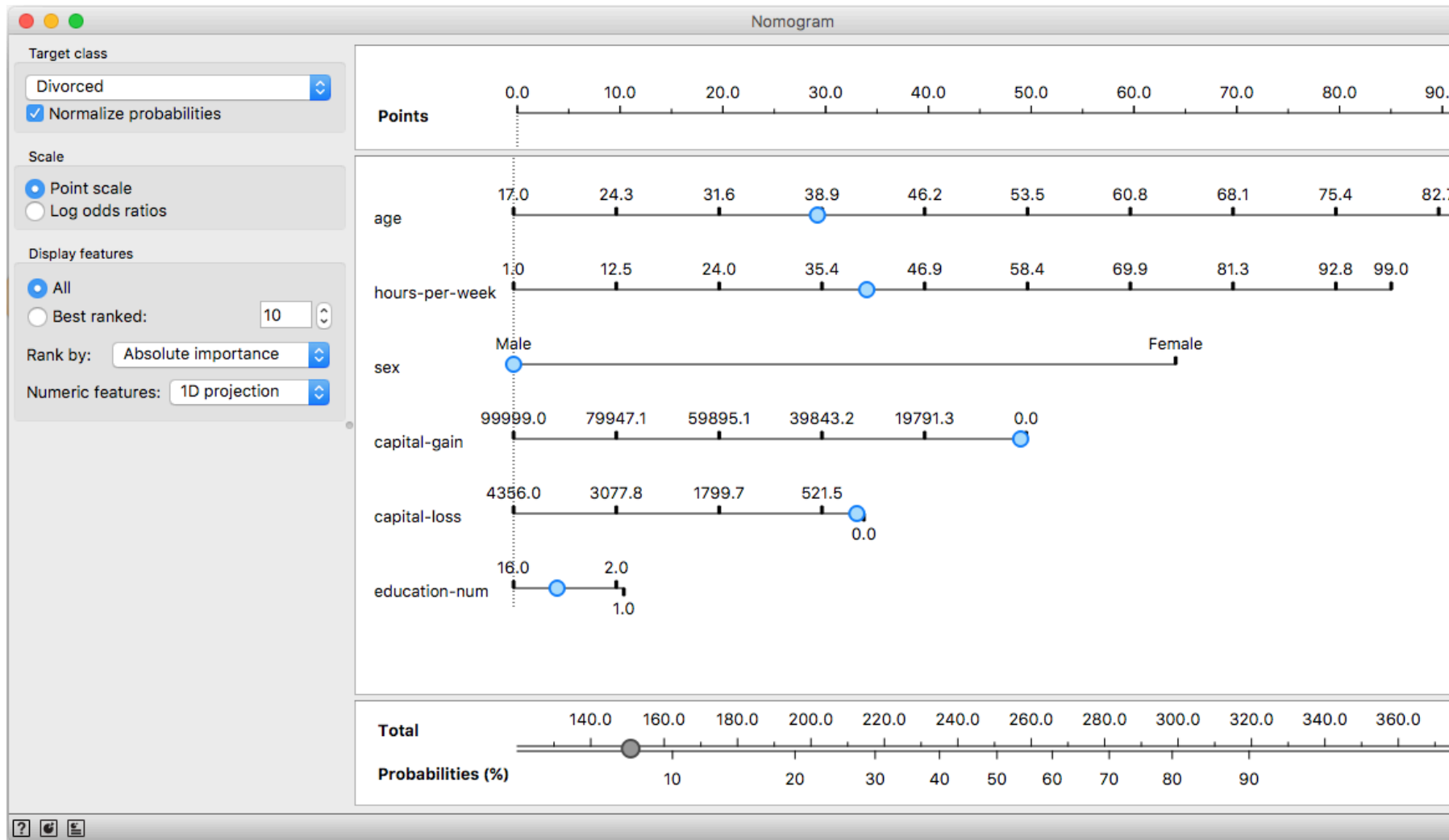
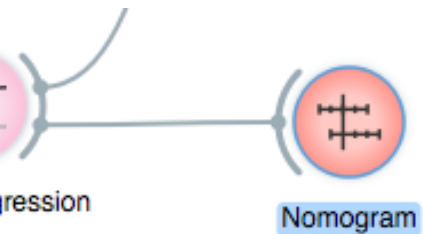
Reset  Send Automatically

Target - Categorical variable



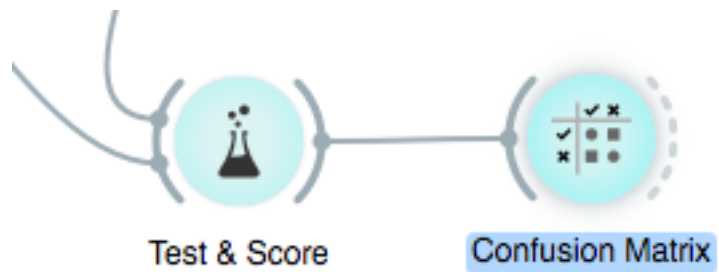
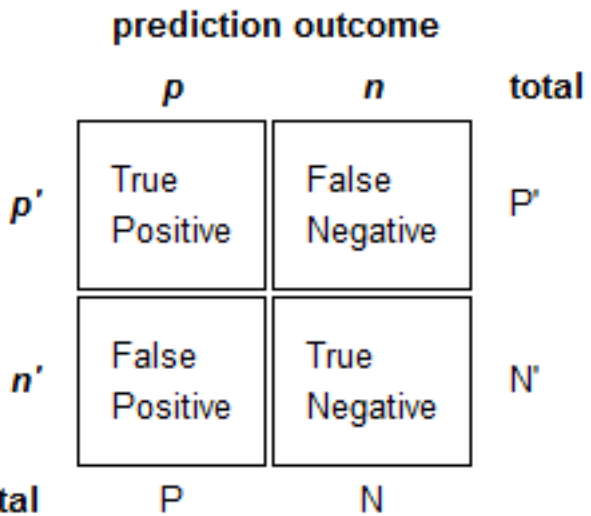
Regression result in nomogram

Adding the points allows to see influence on the variable

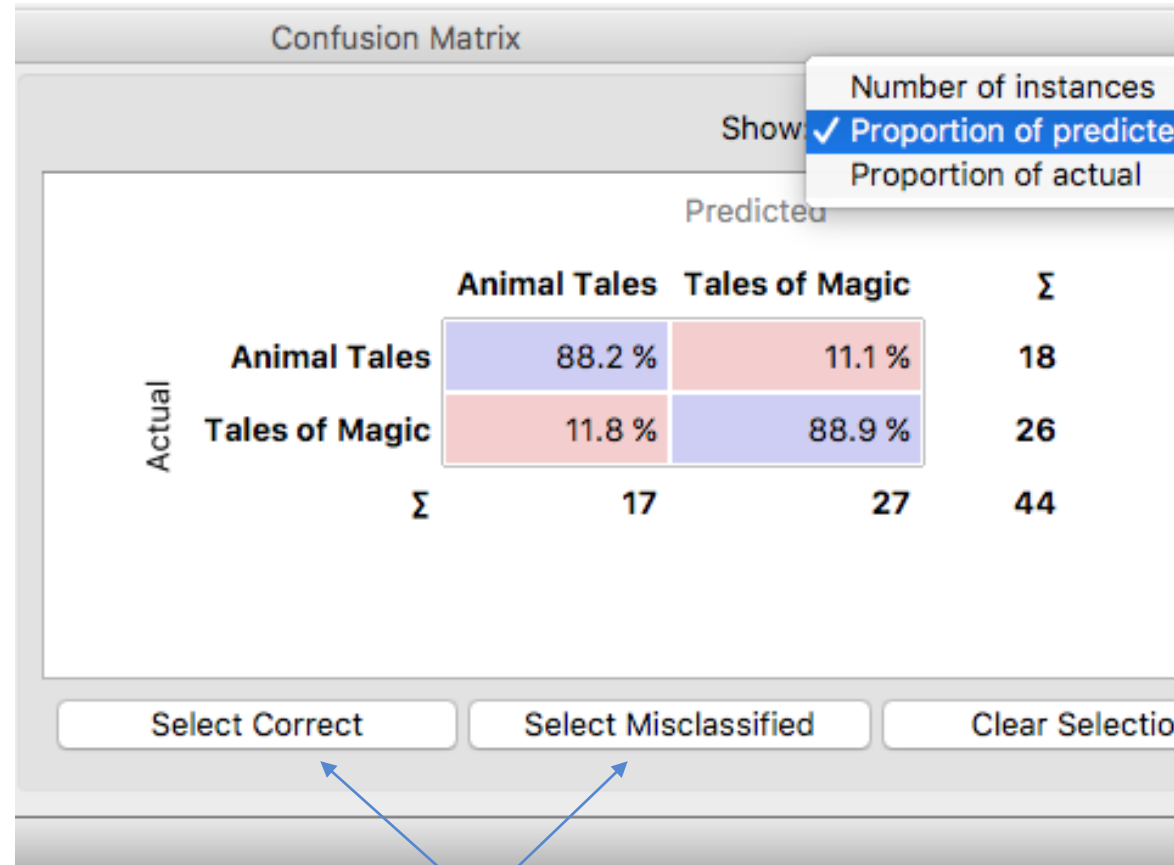




# Confusion Matrix



93.9% tales that are "Animal Tales" identified correctly  
 85.7% tales that are "Tales of Magic" identified correctly

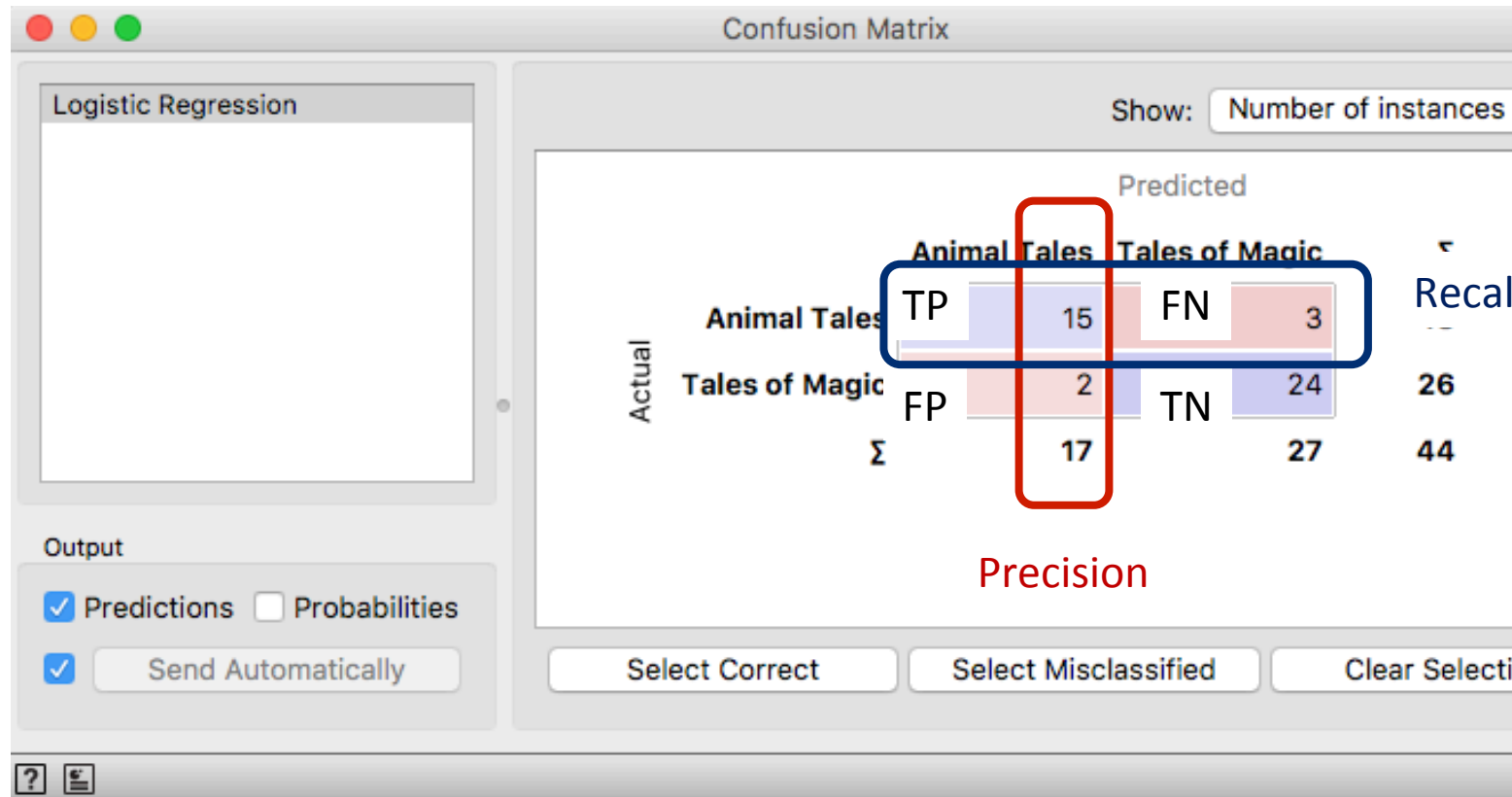


Grey color - selected

class (Animal Tales)- POSITIVE  
d class (Tales of Magic) - NEGATIVE

$$\text{Precision} = \frac{tp}{(tp+fp)}$$

$$\text{Recall} = \frac{tp}{(tp+fn)}$$



Precision for Animal Tales:  $15 / (15 + 2) = 88\%$

Precision for Tales of Magic:  $24 / (24 + 3) = 89\%$

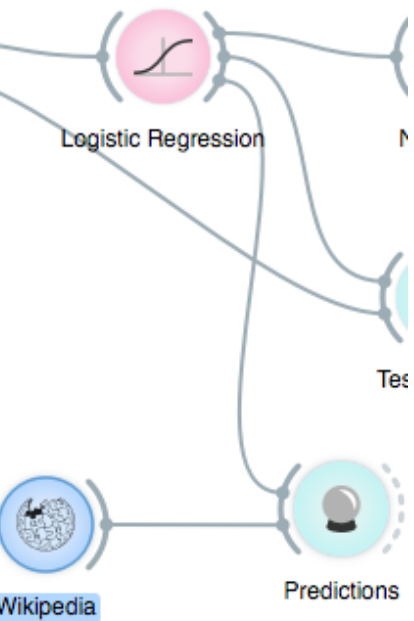
Recall for Animal Tales:  $15 / (15 + 3) = 83\%$

Recall for Tales of Magic:  $24 / (24 + 2) = 92\%$

# We trained the model



Feed new data without labels to the model and ask to classify it.  
Wikipedia search of "aesthetics"



Predictions			
	Logistic Regression	Title	
1	<u>0.38 : 0.62</u> → <u>Tales of Magic</u>	Aesthetics	A
2	<u>0.50 : 0.50</u> → <u>Tales of Magic</u>	Japanese aesthetics	T
3	<u>0.55 : 0.45</u> → <u>Animal Tales</u>	Rasa (aesthetics)	A
4	<u>0.50 : 0.50</u> → <u>Tales of Magic</u>	Practical aesthetics	P
5	<u>0.52 : 0.48</u> → <u>Animal Tales</u>	Aesthetics of music	Ir