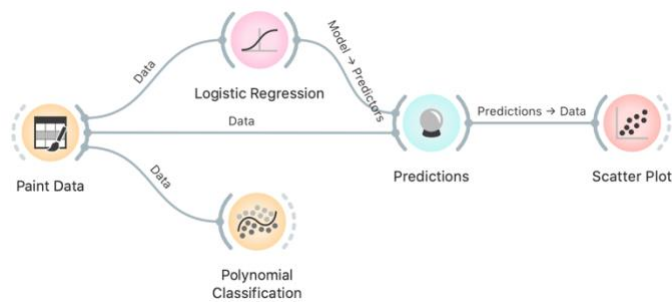
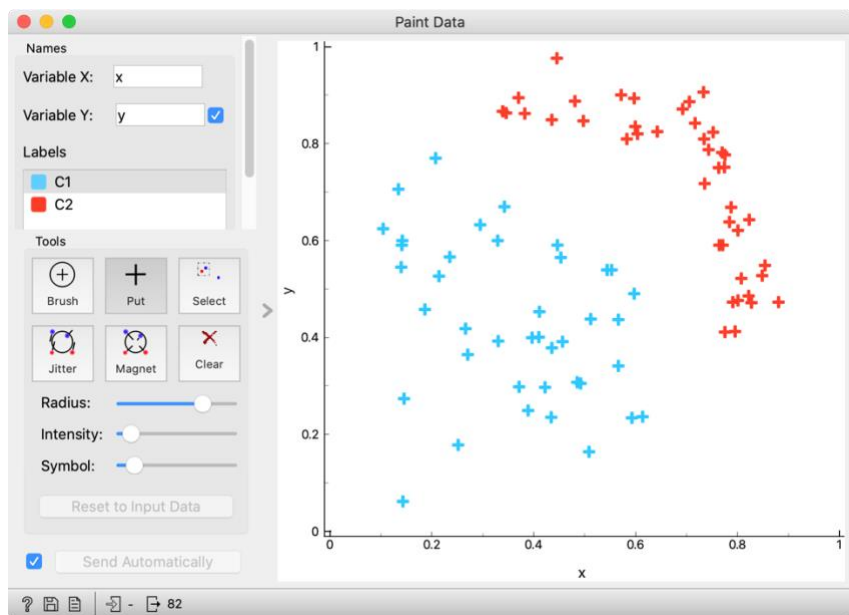


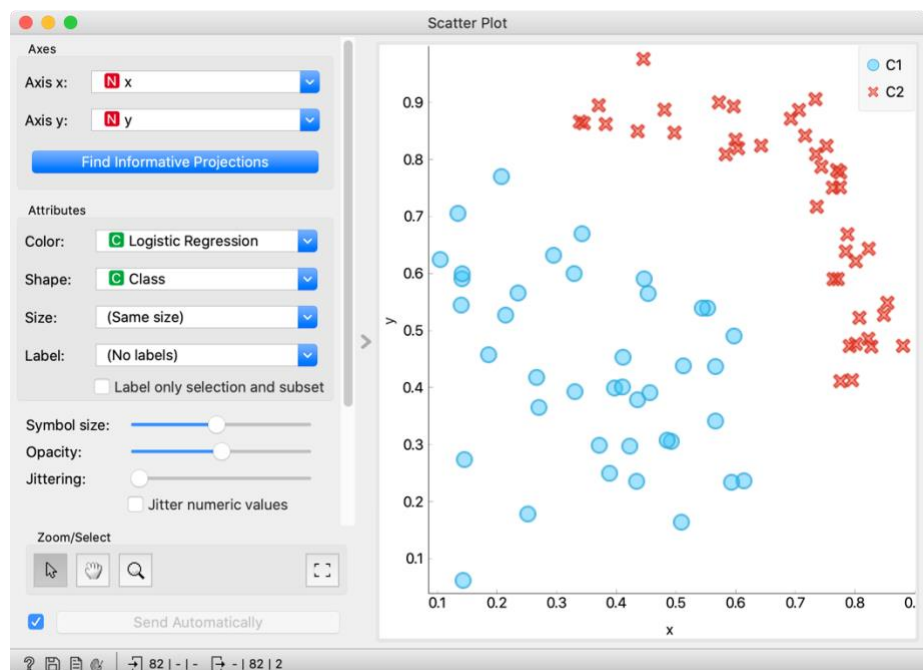
### Homework #3



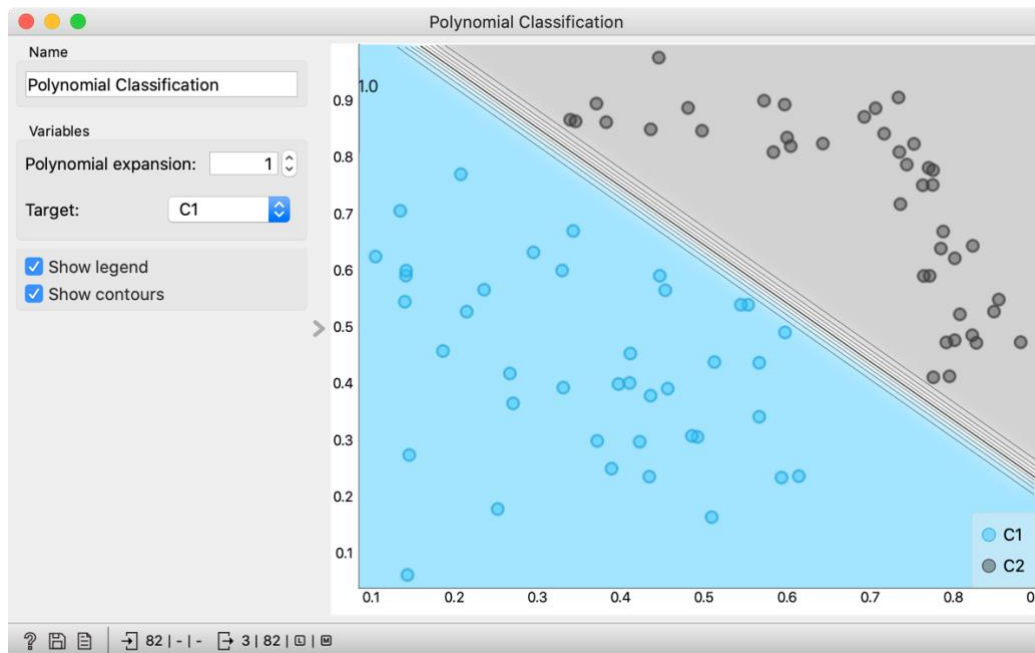
We used the Paint widget to paint some data. We fed the data to Logistic Regression, which produces the model that we gave to Predictions. Predictions also needs some data; we gave it the data we painted. The widget Predictions uses the logistic regression model predict classes for all data instances. We pass the results to the Scatter Plot and set the color and shape to represent the predictions and the actual classes.



We see that colors match the shapes - all circles (actual members of C1) are colored blue (classified as C1) and all crosses are colored red (C2 being classified as C2).



In the Polynomial classification widget, we can see a decision boundary and lines representing probability contours.



In this homework we will explore decision boundaries of different models.

1. Paint some data in which the two classes are clearly separated, yet logistic regression model is unable to separate them. Argue (in a sentence or two) why not.
2. Show some data (it can be the same as the one above), which cannot be separated by logistic regression, but Polynomial regression can do it if we add quadratic terms. Explain why.
3. Now paint some data that requires cubic terms.
4. Observe the behaviour of classification trees on the latter. Argue why they do or do not work.
5. Try to paint some data that can be separated by logistic regression and has enough data instances, yet classification trees would have problems with it – not necessarily being unable to do the job, but being obviously misfit for it. In other words, think about the kind of linearly separable data for which the trees are inappropriate.
6. Support Vector Machines (SVM) are a model that we will mention in our next lecture. At the heart of a model is a "kernel", which allows the model to "simulate" adding quadratic or cubic terms (and worse) without actually doing it. Take the painted data you prepared while answering the previous questions, and observe and comment the behaviour of SVM using
  - Linear Kernel
  - Polynomial Kernel with  $d=2$
  - Polynomial Kernel with  $d=3$
  - RBF kernel

Submit your homework as a report in PDF (not Word!) named lastname-firstname-3.pdf (like demsar-janez-3.pdf; no spaces in the name, lowercase letters, dashes between parts). Number your answers. Due to many images, there is no page limit, but be reasonable.