Semi-supervised learning

Tasks (Lab 14):

1. Generate artificial datasets:

```
X, y = make_circles(n_samples=1000, noise=0.1, factor=0)
```

```
X, y = make_classification(n_samples=1000, n_features=2, n_informative=2)
```

2. Split datasets into training and testing subsets
3. In training data, choose randomly g observations from positive class and g observations from negative class to be labeled. We treat the remaining observations as unlabeled
4. Compare 4 methods semi-supervised learning methods available in scikit-learn library:
   (a) **Naive Method:** learn classifier using only labelled examples in training data.
   (b) **Self-training**
   (c) **Label propagation**
   (d) **Label Spreading**

   The last three methods are based on both labeled and unlabeled data.
5. Use SVC classifier as a base classifier.
6. Compute accuracy on testing data.
7. Analyse how the value of $g = 1, 2, \ldots$ affects the results.
8. Repeat the experiment multiple times and generate box plots showing the accuracy distribution for different values of g.