# Classification of the CIFAR-100 dataset - Report

## Tasks:

- Binary classification into two chosen superclasses.
- Multiclass classification into the fine classes of one superclass.

## Classifiers used:

- K-Nearest-Neighbors (KNN)
- Decision tree
- Multilayer Perceptron (MLP)

## Task 1

For this task the following superclasses of the CIFAR-100 dataset were chosen:

- Aquatic mammals
- Non-insect invertebrates

K-Nearest-Neighbors (KNN) classifier

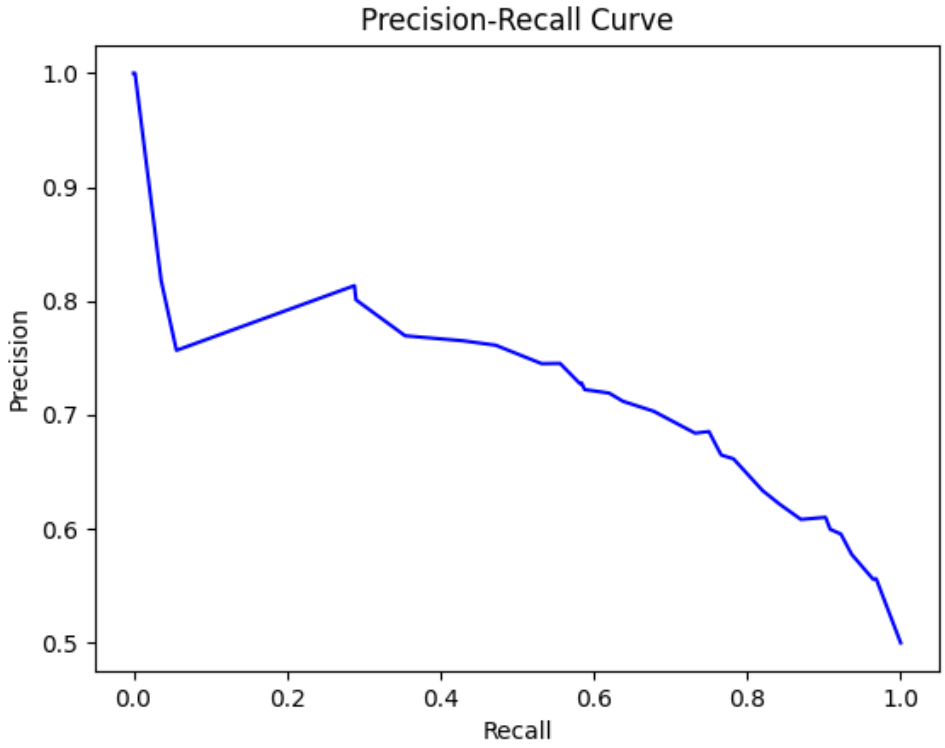
To choose the appropriate k-value for this classifier, cross-validation on this parameter has been performed in the range [1,11).

Results:

- Confusion matrix:

| | |
|---|---|
| 331 | 169 |
| 134 | 366 |

- Macro precision: 0.698
- Recall score: 0.697




Decision tree classifier

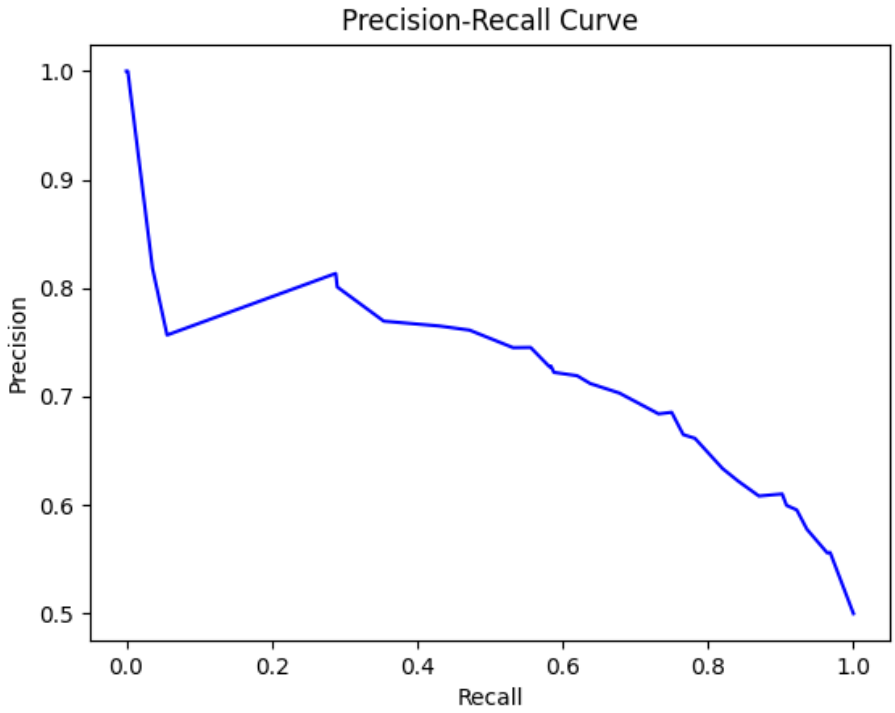
To choose the appropriate maximum depth for this classifier, cross-validation on this parameter has been performed in the range [3,21).

Results:

- Confusion matrix:

| | |
|---|---|
| 331 | 169 |
| 134 | 336 |

- Macro precision: 0.698
- Recall score: 0.698




Multilayer Perceptron (MLP) classifier

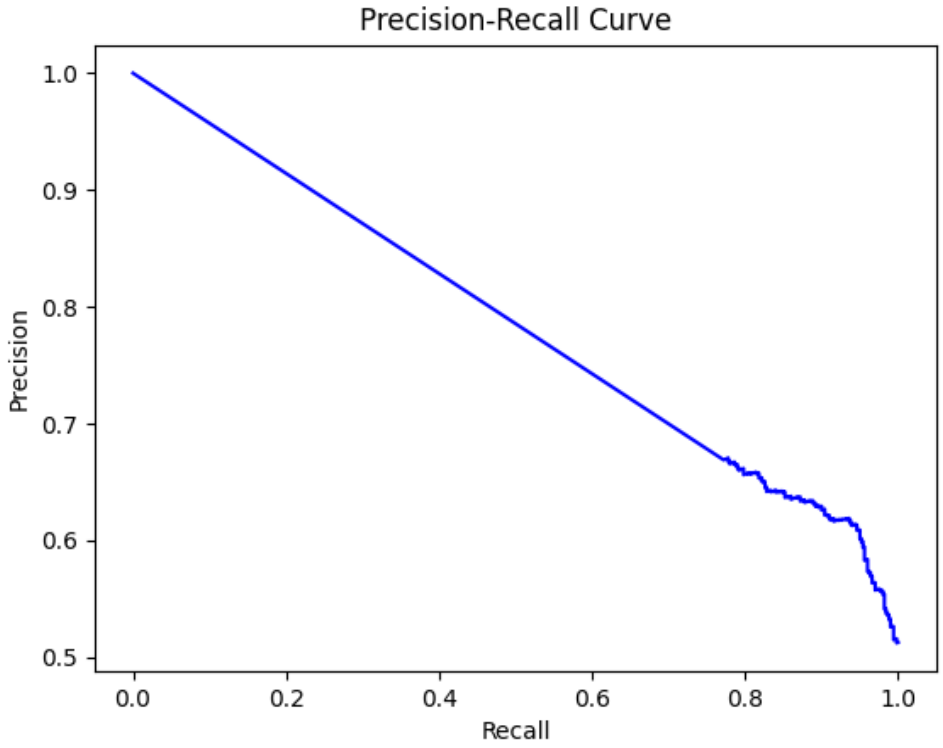
For this classifier no cross-validation was performed, the value for the max_iter parameter is equal to 100.

Results:

- Confusion matrix:

| | |
|---|---|
| 243 | 257 |
| 56 | 444 |

- Macro precision: 0.723
- Recall score: 0.687




## Task 2

For this task the following superclasses of the CIFAR-100 dataset were chosen:

- Bottle
- Bowl
- Can
- Cup
- plate

K-Nearest-Neighbors (KNN) classifier

To reduce the dimensionality on the data, Principle Component Analysis (PCA) was applied to the dataset. We want to choose the number of components that will give us 95% of the variability. Therefore, parameter n_components=0.95.

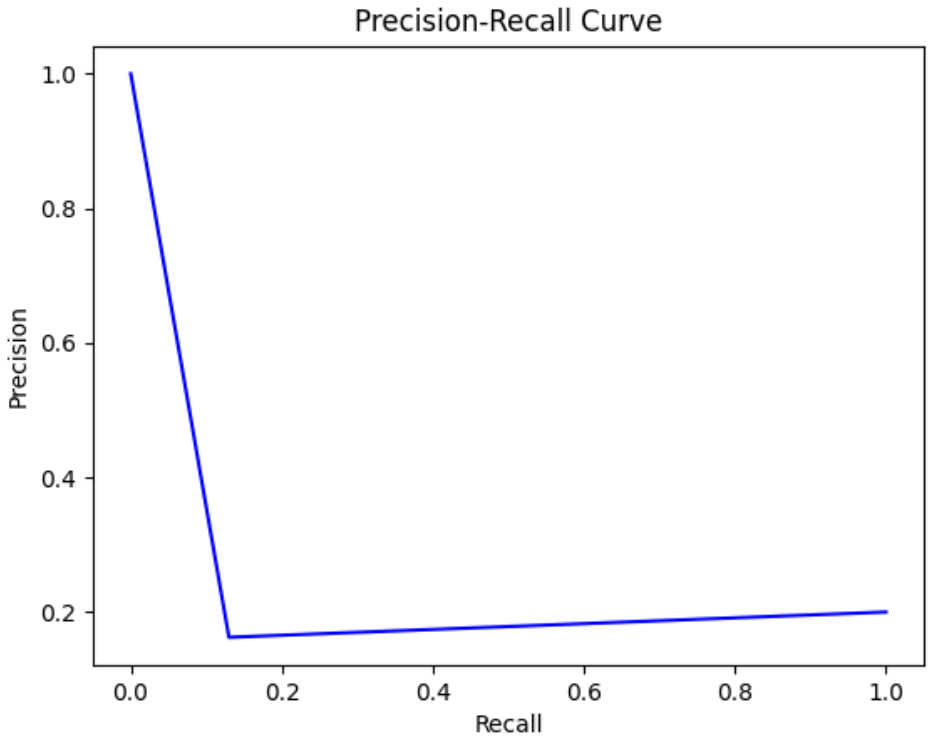
Initially, PCA was not used, however, for KNN (and for the other classifiers from the 2nd task) a boost in performance was seen after applying it.

Results:

- Confusion matrix:

| | | | | |
|---|---|---|---|---|
| 63 | 5 | 8 | 12 | 12 |
| 12 | 39 | 8 | 19 | 22 |
| 13 | 14 | 43 | 20 | 10 |
| 12 | 9 | 4 | 66 | 9 |
| 6 | 13 | 4 | 7 | 70 |

- Macro precision: 0.565
- Recall score: 0.562



Decision tree classifier

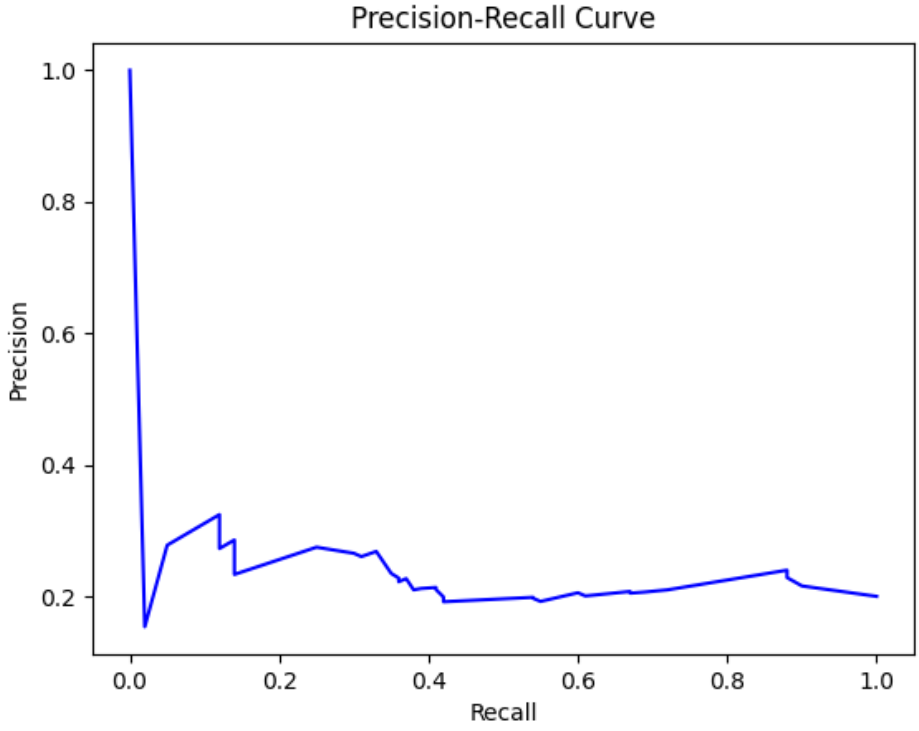
Originally, for this classifier, only cross-validation was used on the max_depth parameter. However, after a very bad evaluation, the performance was boosted by performing grid-search on the criterion parameter (values 'gini' and 'entropy'). Additionally, and in a similar manner to the KNN classifier, PCA was performed on the dataset.

Results:

- Confusion matrix:

| | | | | |
|---|---|---|---|---|
| 44 | 10 | 17 | 10 | 19 |
| 20 | 28 | 19 | 13 | 20 |
| 14 | 15 | 41 | 10 | 20 |
| 9 | 16 | 19 | 41 | 15 |
| 9 | 26 | 8 | 7 | 50 |

- Macro precision: 0.411
- Recall score: 0.408



Multilayer Perceptron (MLP) classifier

As with the previous 2 classifiers, we performed PCA on the dataset here as well. And the max_iter parameter was chosen to be 100.

Results:

- Confusion matrix:

| | | | | |
|---|---|---|---|---|
| 70 | 7 | 9 | 9 | 5 |
| 13 | 55 | 11 | 9 | 12 |
| 16 | 16 | 51 | 10 | 7 |
| 9 | 10 | 3 | 71 | 7 |
| 5 | 24 | 6 | 5 | 60 |

- Macro precision: 0.618
- Recall score: 0.614