Readme

Introduction：

In Q1, we extract feature by ResNet, and then choose the classic outlier detection algorithm to do outlier detection. The detailed information list below.

Methods:

1. Feature Engineering

We use Res-net to extract features with resnet50 pretrained features(dim=2048):

ResNet50(weights='imagenet', include_top=False, pooling='max')

It's a mature neuron network for image detection. On the ImageNet dataset, residual nets with a depth of up to 152 layers, which is useful for feature engineering. As the pre-trained model have a better performance, we use it directly.

2. Outlier detection

2.1 overview of data and local evaluation

Outlier percentage= 0.0792, I assume that the train set and test set is independently and identically distributed, there may exist 100 pictures with labeled "outlier".

And as 2048 features seems large to train, I firstly try to PCA method and reduce the dimension to 500. However, the precision decreases as the component decreased. So, I keep all the features into the models.

2.2 Methods and score

I have trained 3 models with the package pyod, which focuses on outlier detection, including: One-class SVM, PCA, KNN. The best model of those are One-class SVM(OCSVM), but still get low rate of recall in the local validation.

References：

[1] https://github.com/yzhao062/pyod/blob/master/notebooks/Model%20Combination.ipynb

[2] https://scikitlearn.org/stable/modules/generated/sklearn.neighbors.LocalOutlierFactor.html

[3] https://hackernoon.com/one-class-classification-for-images-with-deep-features-be890c43455d