

Layer-wise Relevance Propagation (LRP)
Machine Learning for Health Informatics
Task 3

Dipl. -Ing. Anna Saranti

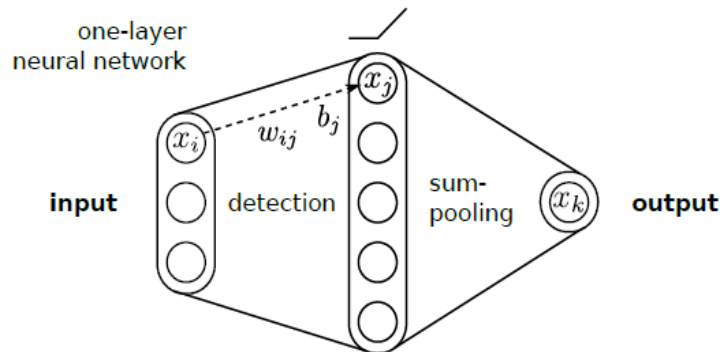
29.05.2020

1 Task description

The task contains two parts:

1. Numerical task

- Use the equations above to compute numerically the relevance of all layers of the network depicted in the figure 1.
- Use your own weight values (w_{ij}), but think on weighting schemes that are typically used in neural networks. See <https://keras.io/initializers/>
- Verify that the conservation and positivity rules properties apply.
- Provide descriptions of the interpretations



2. Programmatic task

- Install Python 3.5.+ and the relevant libraries.
- Provide descriptions of the interpretations of the relevance images with respect to the input images as well as their differences

Python libraries:

- <https://www.tensorflow.org/>, <https://mxnet.apache.org/>
- <https://keras.io/>
- <https://github.com/albermax/innvestigate>

- Run the `examples/readme_code_snippet.py` with any of the `.jpg` figures in the `examples/images` folder (line 37)
- Adapt line 54 to select a different analyzer
- Python IDE: <https://www.jetbrains.com/pycharm/>

2 Literature

NNs:

- Richard O. Duda, David G. Stork, Peter E.Hart "Pattern Classification" Chapter 6.
<https://github.com/dazzz/patrec2015/blob/master/Pattern%20Classification%20by%20Richard%20Duda%20David%20Stork%20Peter%20E.Hart%20.pdf>
- Christopher M. Bishop "Pattern Recognition and Machine Learning"
<http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf>

LRP:

- Montavon, Grégoire, et al. "Explaining nonlinear classification decisions with deep taylor decomposition." *Pattern Recognition* 65 (2017): 211-222.

3 Questions

- Anna Saranti:anna.saranti@medunigraz.at