**Antwoorden assignment 3**

**Exercise 1**

The PatchCAMELYON challenge on Kaggle uses the area under the ROC curve as an evaluation measure. Describe the concept of ROC curve analysis and the area under the ROC curve (AUC) as an evaluation measure. Then, using methods from the sklearn Python toolbox (the required functions are alerady imported), perform ROC curve analysis and computation of the AUC for your model.

Note that you will have to do this analysis on the validation set (since you do not have access to the ground truth for the test set).

*When 1-specificity (i.e. False Positive Rate, FPR) on the x-axis is plotted against sensitivity (True Positive Rate, TPR) on the y-axis, the plot is called the Receiver Operator Characteristic (ROC) curve [A]. To compare different classifiers, i.e. the same classifier with different learning parameters or completely different classifiers, the ROC curve can be useful [boek]. A single run of one classifier gives a single point on the ROC plot [boek].*

*In medical context for instance, the basic principle of the ROC curve is to quantify how accurately medical diagnostic tests can discriminate between two patient states e.g. diseased or healthy [A]. By altering the decision threshold we will get different fractions of TP and FP [B]. For all these decision thresholds then the fractions are plotted in the ROC curve. A classifier is deemed to be 'perfect' when it is located at the point (0,1) on the ROC curve, i.e. when the classifier has a TPR of 1 and a FPR of 0 for a specific decision threshold. Moreover, when the ROC curve corresponds to the 45 degree line (y = x), the diagnostic test that belongs to this curve is as good as random guessing (i.e. chance level) [A]. In other words, we have a test which yields positive or negative results unrelated to the true disease status [A]. Hence, the closer to the top-left-hand corner the result of a classifier is, the better it has performed [boek]. Thus, to compare different classifiers, one can use the Area Under the Curve (AUC) to evaluate and compare model performance. Since the closer the classifier gets to the 'ideal' point of (0,1), the larger the AUC and the better the classifier performs [boek].*

[A] Hajian-Tilaki K. (2013). Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. Caspian journal of internal medicine, 4(2), 627–635. [B] Charles E. Metz, Basic principles of ROC analysis, Seminars in Nuclear Medicine, Volume 8, Issue 4, 1978, Pages 283-298, ISSN 0001-2998, <https://doi.org/10.1016/S0001-2998(78)80014-2>. (<https://www.sciencedirect.com/science/article/pii/S0001299878800142>) [boek] Marsland, S. (2014). Machine Learning: An Algorithmic Perspective, Second Edition (Chapman & Hall/CRC Machine Learning & Pattern Recognition) (2de editie). Chapman and Hall/CRC.

## Exercise 2

It is possible to construct a neural network model that is equivalent to the model above, however, only using convolutional layers (i.e. without using any fully connected or "dense" layers). Construct and train such a model.

What would be the advantage of only using convolutional layers?

Equivalent? Do they mean using a convolutional neural network with a kernel of 1 ? But what would be the advantage of that? Dense can be replaced by convolutional layer with kernel 1 and then we need to replace max pooling layer by convolutional layer?

It is possible to construct a neural network model using only convolutional layers. When using multiple convolutional layers, each convolutional layer is assigned their own features to detect. Usually, the first convolutional layer is used for low-level feature detection, e.g. edges, color, gradient orientation etc. [C]. The next layers then combine these features into more complex features to allow for the detection of specific characteristics of an image.

The architecture of a convolutional neural network, in comparison to an ordinary feed forward neural network, allows for better fitting to the image dataset caused by a reduction in the number of parameters involved and the reusability of weights [C]. Using only convolutional layers drastically reduces the amount of weights that needs to be calculated and updated during back propagation. An advantage is thus that the computation time is decreased and overall the creation of the neural network model is less computationally expensive. In summary, the goal of a convolutional neural network is to reduce the images into a form which allows for easy processing without losing features that are necessary for getting an accurate prediction [C].

[C] <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>