Deep Layers

WORKSHOP / ARTIFICIAL INTELLIGENCE

20. - 21. 9 . 2022

Institute of Scientific Instruments of the Czech Academy of Sciences

Kralovopolska 147, Brno, Czech Republic

Registration (free, but mandatory) | www.isibrno.cz/deep

Deep Layers a machine-learning forum

STRATEGIE AV21

brno ai

AI MT

Průlomové technologie budoucnosti

ISI CAS

Machine Learning Elements II – Model training

How to devide, preprocess data and build basic ML models

Filip Plesinger

ISI of the CAS, Brno, CZ

You are welcome to experiment with dataset during the lesson.

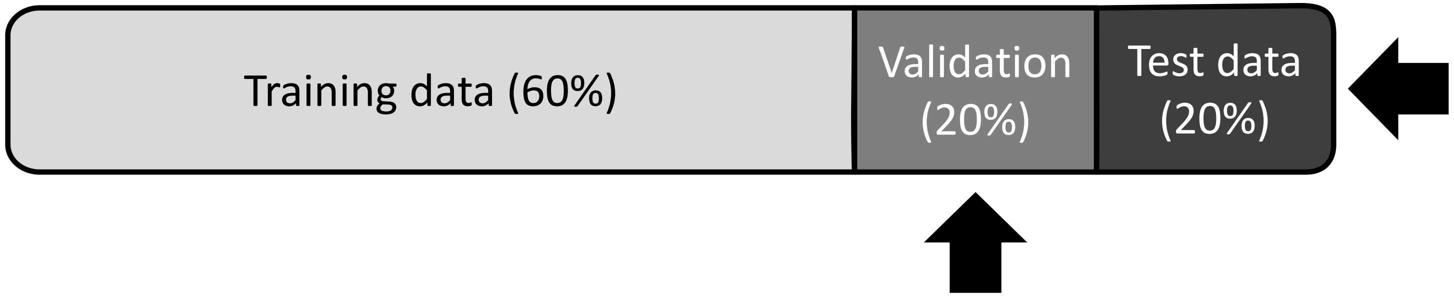
QR code link to COLAB NOTEBOOK :
(Or through https://www.isibrno.cz/deep/)

1. Splitting data

Link to COLAB NOTEBOOK:

Split to training & testing data:

More correctly (and definitely, for DL) it should be:



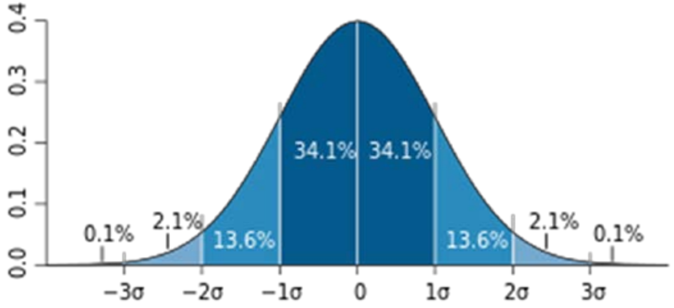
For final testing (only once)

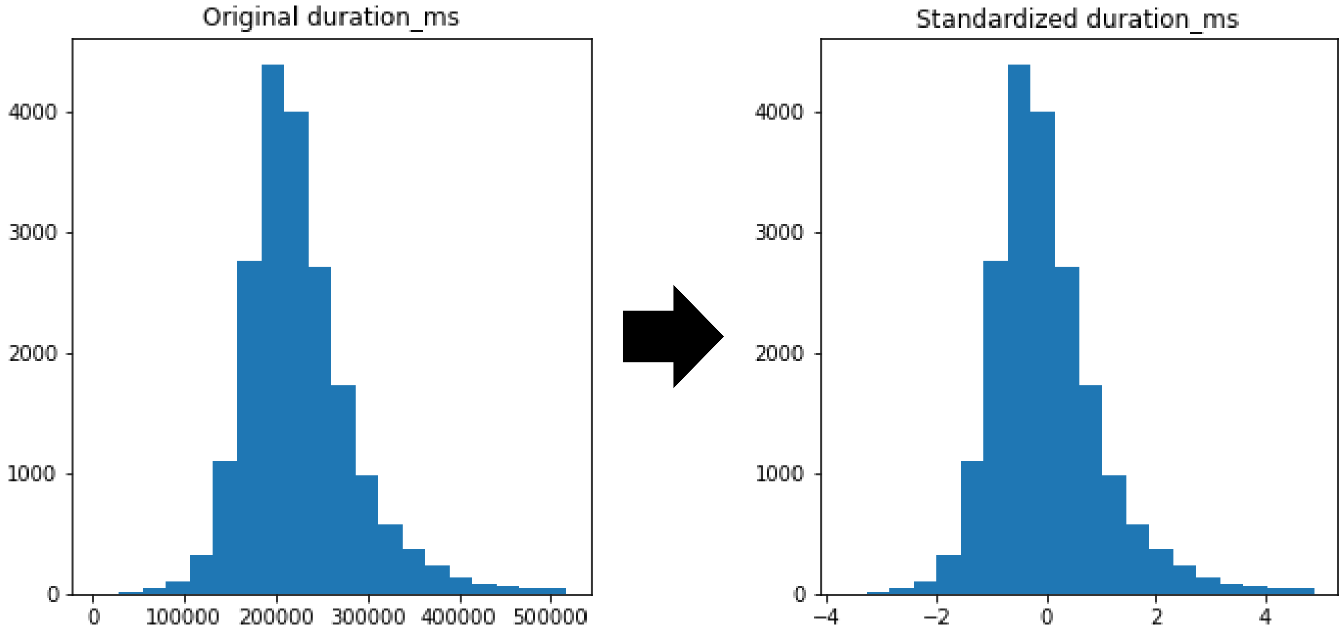
For hyper.parameter tuning, feature selectiom, model selection

2. Bring into consistent scale

Standardization:

$$x_{std}^{i} = \frac{x^{i} - \mu_x}{\sigma_x}$$





Required by most ML/DL methods

3. Logistic regression

- Prefrectly explanable
- Easy implementation
- Weaker performance

$$y = \frac{1}{1 - e^{-(i + a\cdot X + b\cdot Y + \ldots)}}$$

Output probability

const	-0.0090
danceability	0.0996
key	0.0204
loudness	0.1595
mode	-0.0080
speechiness	-0.0521
acousticness	0.1412
instrumentalness	-0.1911
liveness	-0.0657
tempo	0.0306
duration_ms	-0.1866

More descriptive form: **statsmodels** package

4. Model performance metrics (classification)

True positive (TP) model predicts 1(popular song) and reality is 1(popular song)
True negative (TN) model predicts 0(less popular) and reality is 0(less popular)
False positive (FP) model predicts 1 and reality is 0
False negative (FN) model predicts 0 and reality is 1

- Default metric in **sklearn** : Accuracy $= \frac{TP+TN}{TP+TN+FP+FN}$
- If the output is **not balanced**, F1 score should be used:

$$\text{F1} = \frac{2 \cdot TP}{2 \cdot TP+FP+FN}$$

Example of an inbalanced dataset:
100 patients, 10 ill, 90 healthy
Classifier says „everybody is healthy“

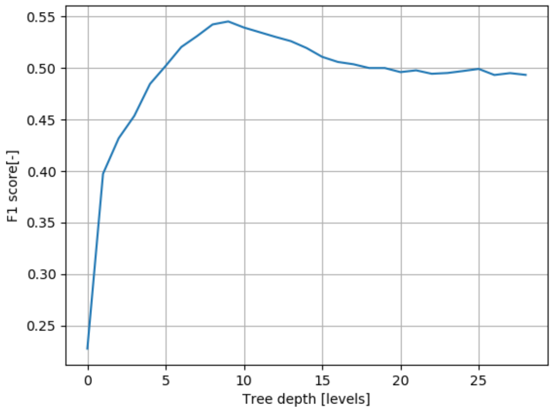
TP	TN	FP	FN	Acc	F1
0	90	0	10	0.9	0.0

Perfect summarization table: https://en.wikipedia.org/wiki/F-score

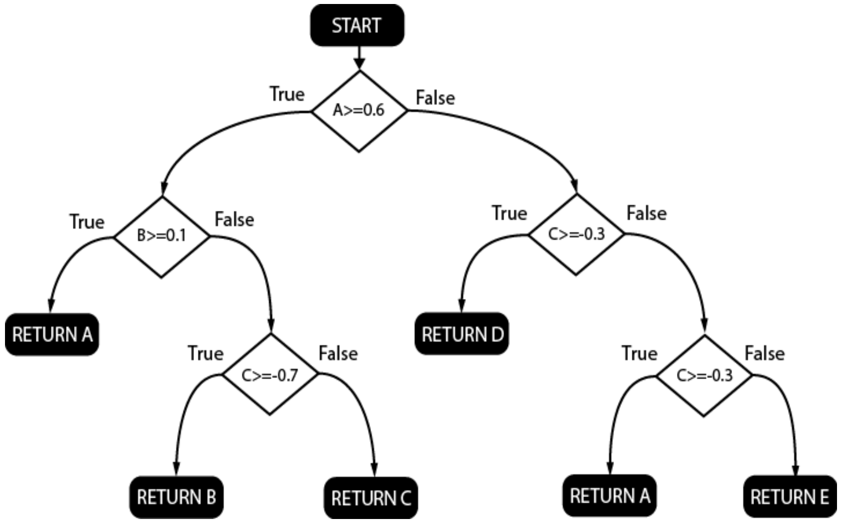
Deep Layers
a machine-learning forum

5. Decision tree

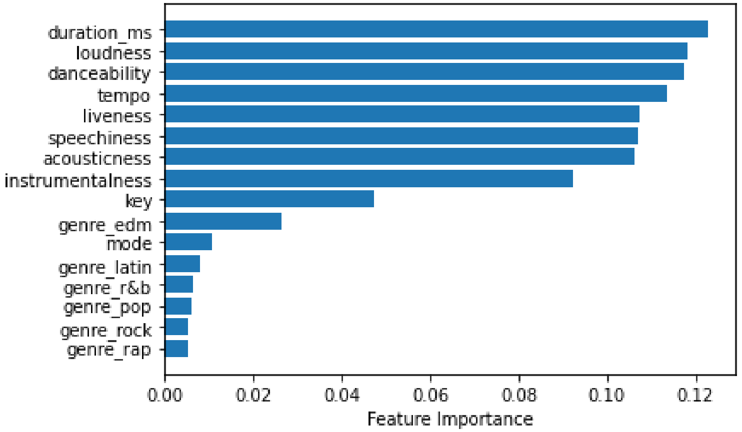
- Explainable
- Easy implementation. Does not care about scale/distribution
- Weaker performance. Easy to overfit



Performance by complexity



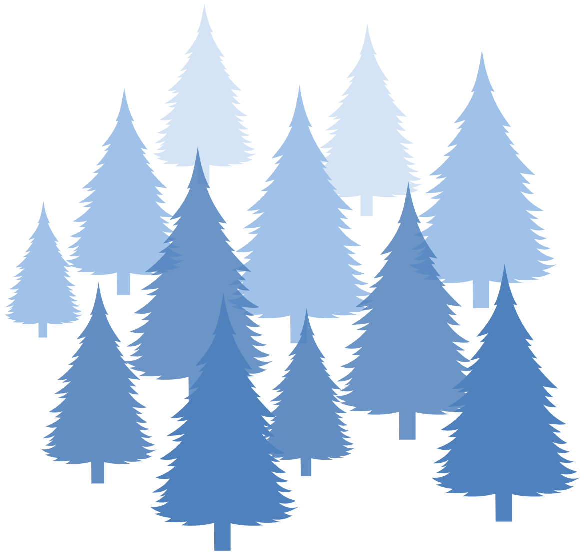
Decision tree example, classifictiom into 6 classes

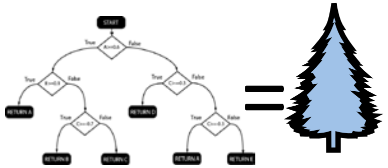


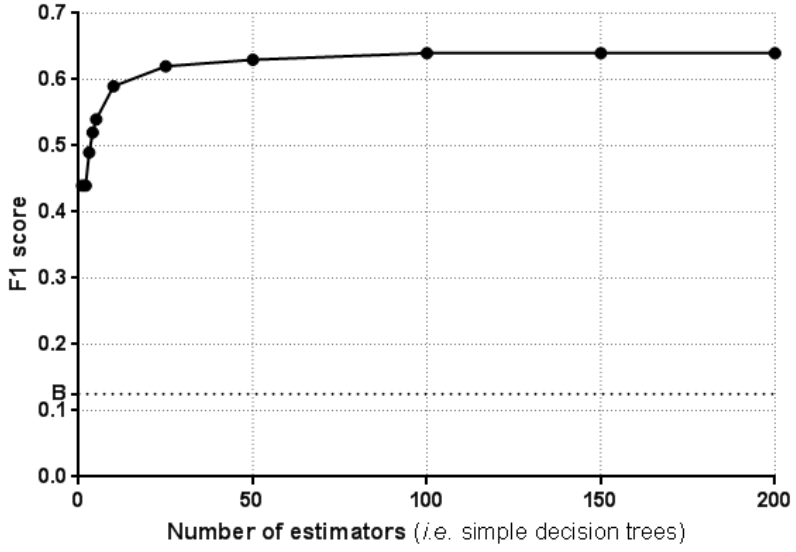
Feature importances in our MUSIC calssification task

6. Random forest

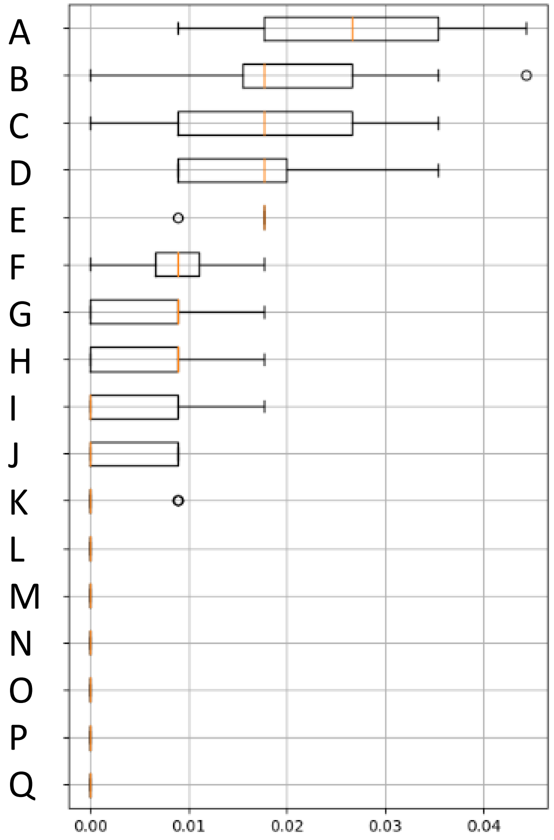
- Still explainable (feature importance only)
- Does not care about … anything







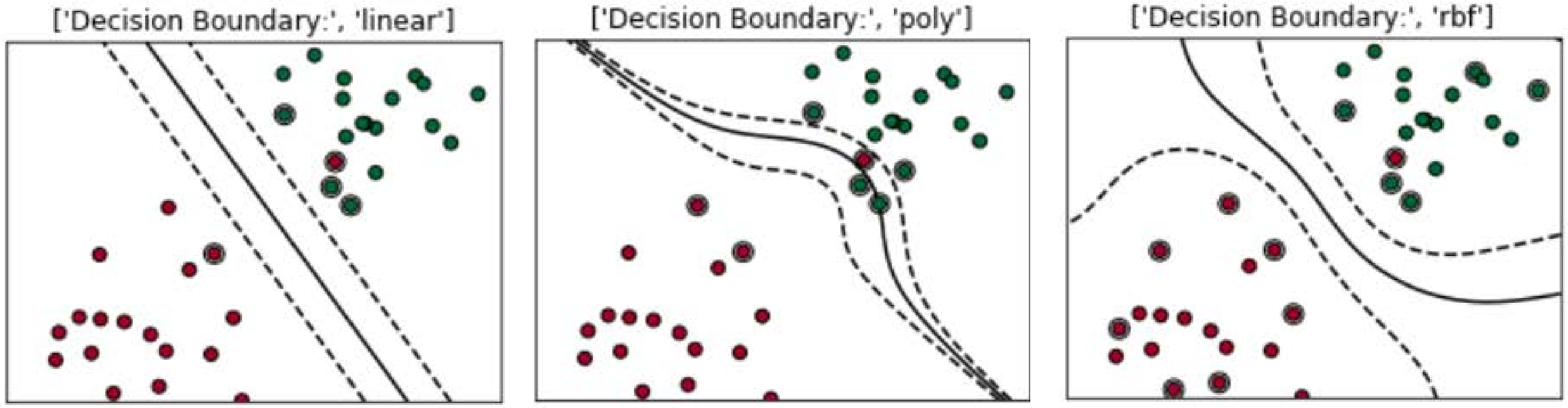
Forest performance by number of trees



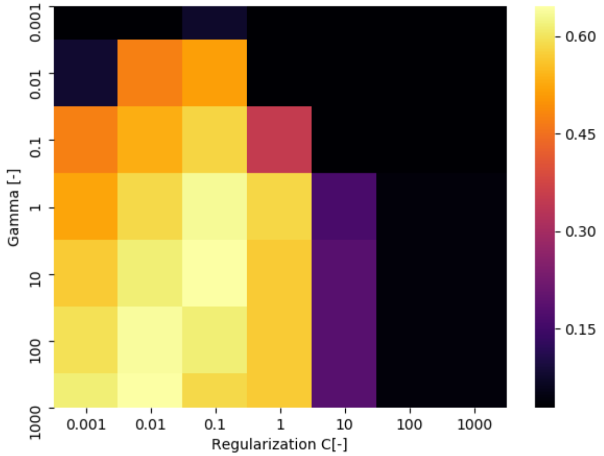
Permutation performance

7. Support-vector-machines (SVM)

- Can use different „kernels“ – linear/polynomial/**radial-basis-function** (RBF) = default
- Stronger performance, but longer training time (the worst from class. ML methods)
- SVMs benefits from hyperparameter optimization



[Source: https://towardsdatascience.com/support-vector-machine-simply-explained-fee28eba5496](https://towardsdatascience.com/support-vector-machine-simply-explained-fee28eba5496)

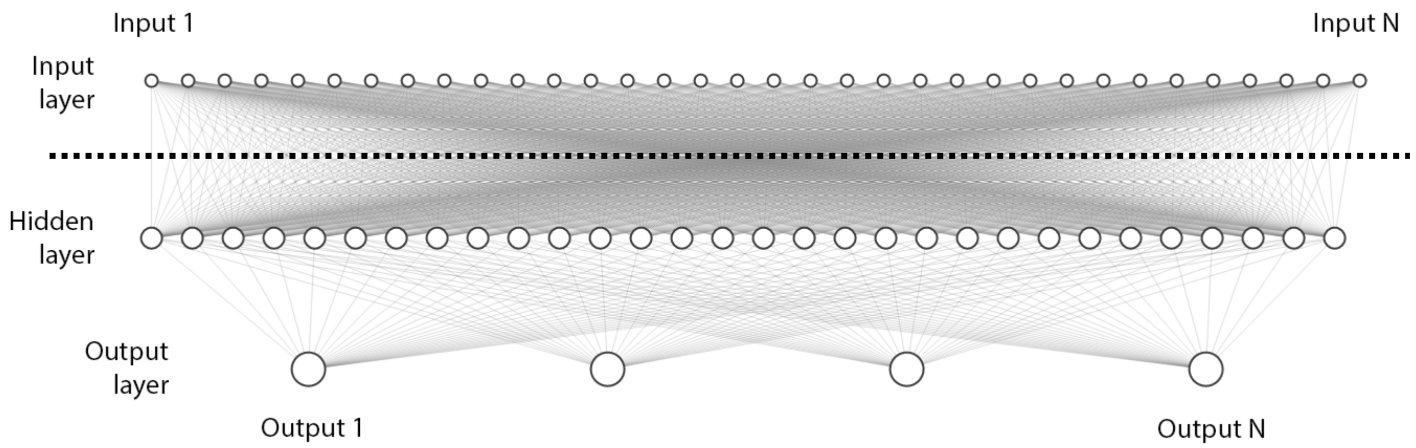


Performance by hyperparameters Gamma & regularization (def. 1)

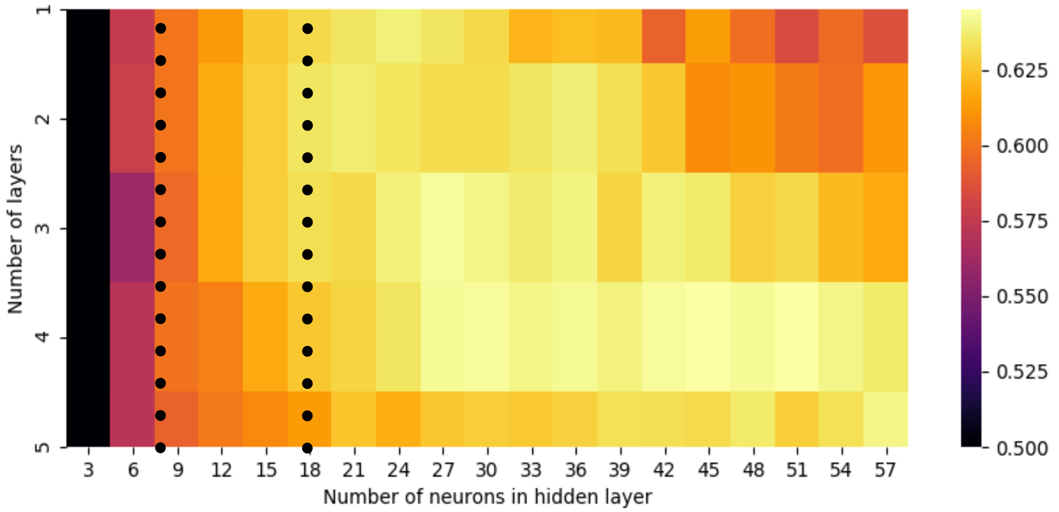
SVM usage in Epilepsy research: tomorrow 11:20 by B. Chybovski

8. Neural network (simple)

- Stronger performance (but usually do not performs better than RF or SVM-RBF)
- Needs „hyperparameter care“, trains longer than other ML methods
- They usually form **final building block of DL networks** (i.e. fully connected layer)



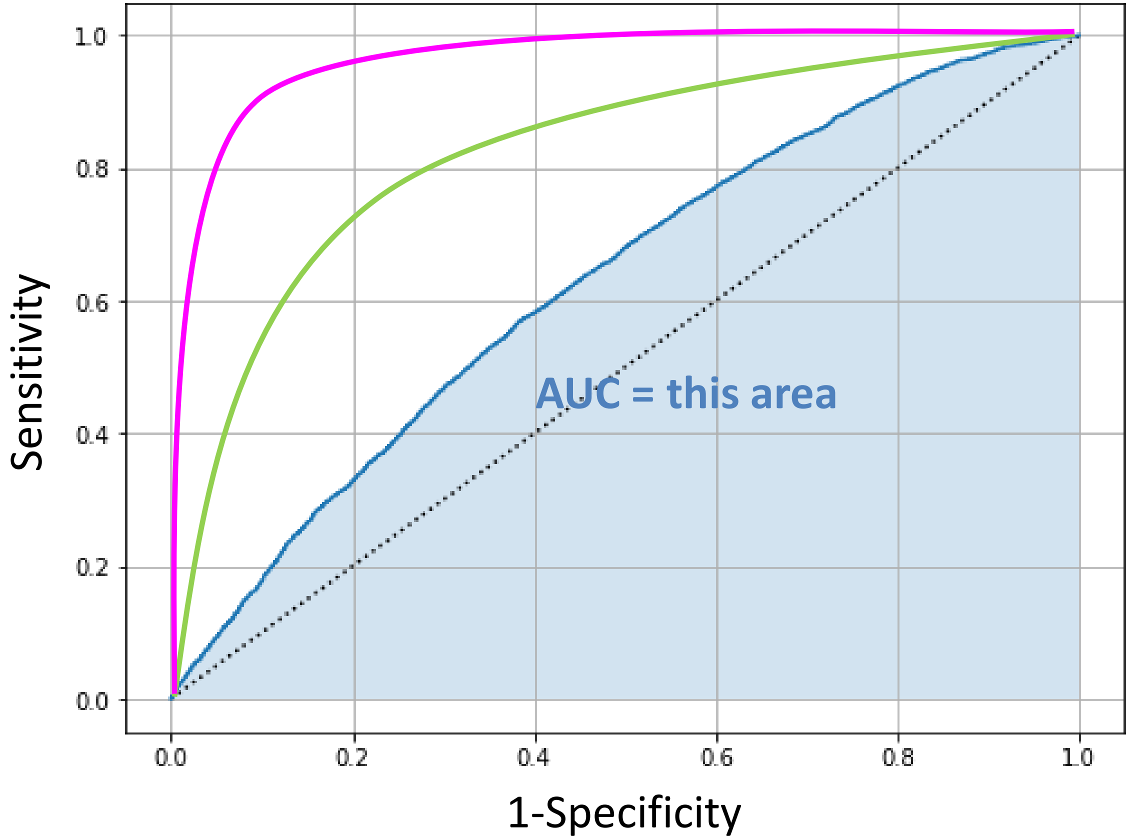
A typical shallow neural network



Performance (color) by neuron count in each hiddden layer and by hidden layer count

9. Comparing models with AUC

- Area under ROC curve (ROC=Receiver-operating-characteristic)



TPR = Sensitivity= $\frac{TP}{TP+FN}$

FPR = 1-Specificity $= 1 - \frac{TN}{TN+FP}$

0.50 – useless classifier

0.63 – better classifier

0.80 – much better classifier

0.90 – wonderful classifier

1.00 – ultimate classifier

<0.50 – probably mismatched labels

10. Summarization

- Incorrect (or none) data split to train/validation/test => incorrect <u>predictive</u> model
- Understanding data behavior is important for **model selection**
- Understanding **the target application** is important for **model selection**
- Different model types require specific data treatment (i.e., careful **feature selection** for LR models)
- It is practical to **standardize** data (but tree-based approaches do not need it)
- Models usually benefit from **hyperparameter tuning** (NNs, SVMs, simple trees)
- **More complex** model **does not necessarily mean** a **better** model performance
- Feature **explainability** depends on model type
- For building mentioned models, you need only the **sklearn** package

Thank you for your attention

Filip Plesinger (fplesinger@isibrno.cz)

Do you have any questions?

Our further activities:

5.10.2022 – ICRC Academy (15:00, here)
Umělá inteligence pro analýzu poruch srdeční činnosti
https://akademie.fnusa.cz/?p=1311

8.11.2022 – SignalPlant workshop (the whole day, here)
signal analysis and processing
www.signalplant.org