sheldon

smart habitat
for the elderly

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

Funded by the H2020 Framework Programme
of the European Union

This publication is based upon work from COST Action CA16226: Indoor Living Space Improvement: Smart Habitat for the Elderly, supported by COST (European Cooperation in Science and Technology).

COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation.

www.**cost**.eu

www.**sheld-on**.eu

Course:

Explainable AI in AAL

Lecture 1:

Explainable AI

Black Box Models

- Most of the models used in AI are black box models
  - Hard to know whether the model can be trusted
  - It's unclear why the model gives the predictions it does
  - Harder to find potential errors
  - Harder to improve model
- While we can understand *how* a model works, it's not easy to tell why the prediction is what it is
- We don't know how the features influence the prediction
  - Prediction could be trivial – e.g., a model with an ID feature which is different for every sample can base the prediction only on the ID and achieve a 100% accuracy on the training set, but it's useless on unknown data
  - Model could incorporate human bias – due to biased data, feature selection

Black Box Models

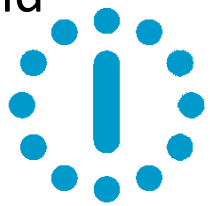
- Some models are more explainable and easier to understand than others
  - Decision trees clearly show how we get to the prediction
  - In linear regression, it's easy to see how the features influence the prediction
- Some models are much harder to understand
  - Neural networks are among the least explainable models – it's not clear why the model returns the prediction it does or how the different features influence this prediction

Need for Explainable AI

- Artificial Intelligence in AAL is focused on machine learning
- Tasks are focused on activity recognition, fall detection, behavioral change and anomaly detection, cognitive impairment detection
  - Important systems with costly mistakes
- Wrong predictions can lead to consequences for the health of the users
  - Important to make sure model is as reliable as possible
  - Important to make sure mistakes are as harmless as possible

Benefits of Explainability

- Easier debugging of misclassifications
  - Why is the prediction wrong?
  - Find out when prediction is wrong
  - How can the model or the training data be improved to reduce wrong predictions?
    - Change model hyperparameters, change connections or layers sizes in neural networks, change training approach, choose new model
    - Feature selection, scaling, choice of more balanced classes, more balanced samples in training features
- Overall model improvement
  - Why is the prediction right? Is it correct for the right reasons?
- Gain new knowledge
  - Explainable AI can reveal new physical, chemical and biological mechanisms (e.g. find genes linked to certain diseases, find new symptoms, etc.)



sheldon

smart habitat
for the elderly

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

Funded by the H2020 Framework Programme
of the European Union

This publication is based upon work from COST Action CA16226: Indoor Living Space Improvement: Smart Habitat for the Elderly, supported by COST (European Cooperation in Science and Technology).

COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation.

www.**cost**.eu

www.**sheld-on**.eu