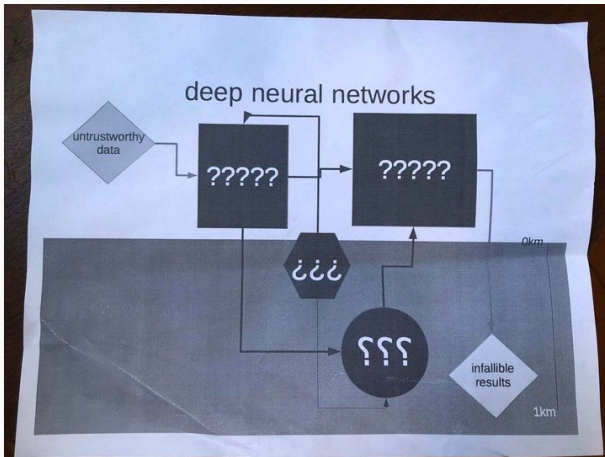


Detecting individual-level deviations in brain morphology with Layerwise Relevance Propagation

Esten Høyland Leonardsen

June 22, 2022

Explainable AI: Motivation



Explainable AI: Motivation

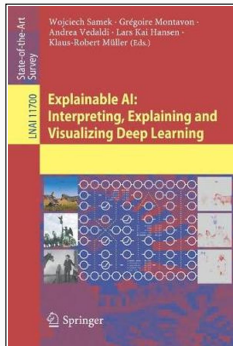


Explainable AI: Motivation

```
Traceback (most recent call last):  
  File "<stdin>", line 1, in <module>  
  File "<stdin>", line 2, in print_model  
  File "<stdin>", line 4, in print_layer  
  File "<stdin>", line 4, in print_layer  
  File "<stdin>", line 4, in print_layer  
  [Previous line repeated 8 more times]  
  File "<stdin>", line 3, in print_layer  
MemoryError
```

“Relying on devices whose logic is opaque violates principles of medical ethics.”

Explainable AI: Overview



Explainable AI: Overview

Recipient

End user
Domain expert
Policy makers
Society
Developer

Level

Explain the model as a whole
Explain single decisions
Produce stereotypical examples

Explainable AI: Overview

Recipient

End user

Domain expert

Policy makers

Society

Developer

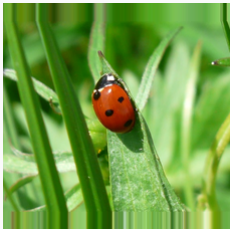
Level

Explain the model as a whole

Explain single decisions

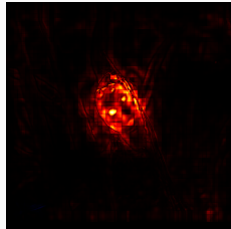
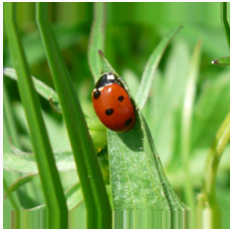
Produce stereotypical examples

Explainable AI: Overview

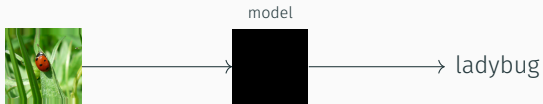


This is a
ladybug because
of the red
back with
the black dots

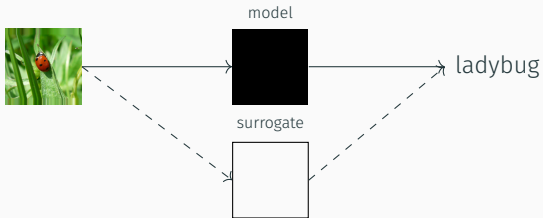
Explainable AI: Overview



Explainable AI: Surrogate models

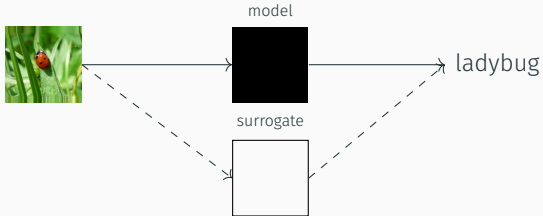


Explainable AI: Surrogate models

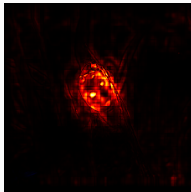


$$y \sim X_1W_1 + X_2W_2 + \dots + X_nW_n$$

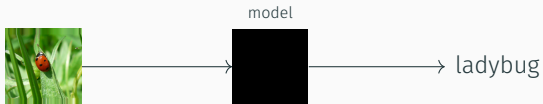
Explainable AI: Surrogate models



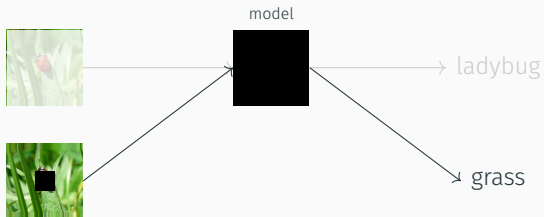
$$y \sim x_1 w_1 + x_2 w_2 + \dots + x_n w_n$$



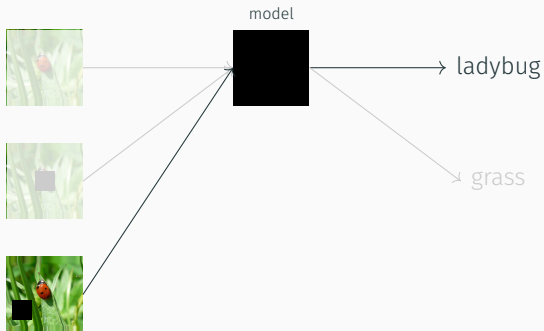
Explainable AI: Occlusion



Explainable AI: Occlusion



Explainable AI: Occlusion



Explainable AI: Occlusion

