

Neural Additive Models: Interpretable Machine Learning with Neural Nets

Edoardo Sarri

Deep Learning Applications

Maggio 2025



images/0-varie/logo UniFi.pdf

Introduzione

XAI

La XAI ha l'obiettivo di spiegare le decisioni prese dai modelli di AI:

- Interpretabile
- Spiegabile

images/0-varie/logo UniFi.pdf

Introduzione

XAI

La XAI ha l'obiettivo di spiegare le decisioni prese dai modelli di AI:

- Interpretabile
- Spiegabile

images/0-varie/logo UniFi.pdf

Modelli lineari

$$f(x) = \sum_{i=1}^D \beta_i x_i + \beta_0$$

- Interpretabile
- Mondo non lineare

GAMs

Modelli lineari

$$f(x) = \sum_{i=1}^D \beta_i x_i + \beta_0$$

Definizione

$$\sigma(y) = \sum_{i=1}^D f_i(x_i) + \beta$$

- f_i è detta shape function.
- σ è detta link function.

Output

L'output è ottenuto in modo additivo.

- Se la link function è l'identità allora abbiamo una regressione.
- Se la link function è una sigmoide allora abbiamo una classificazione.

Neural Additive Models: Interpretable Machine Learning with Neural Nets

Edoardo Sarri

Grazie



images/0-varie/logo UniFi.pdf

Temporary page!

\LaTeX was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it. If you rerun the document (without altering it) this surplus page will go away, because \LaTeX now knows how many pages to expect for this document.