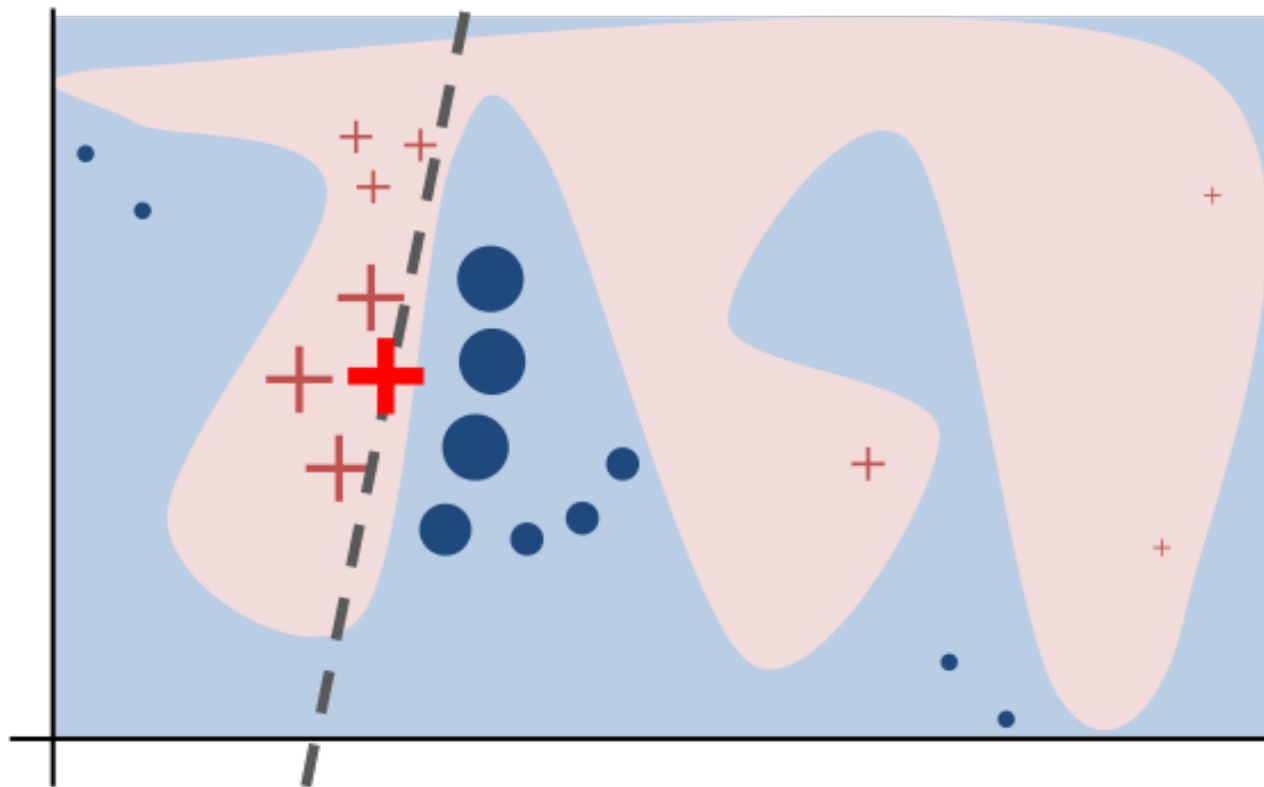


Interpretability
through Explanations



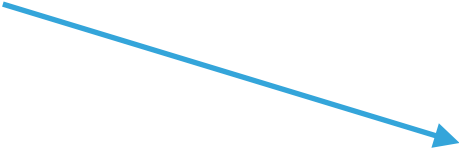
- Perturbations [Ribbeiret al, '16; Lundberg & Lee, '17]; ...

[Image credit: Ribbiero et al.]

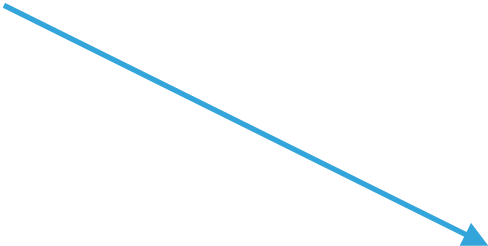




Complex model's
decision boundary



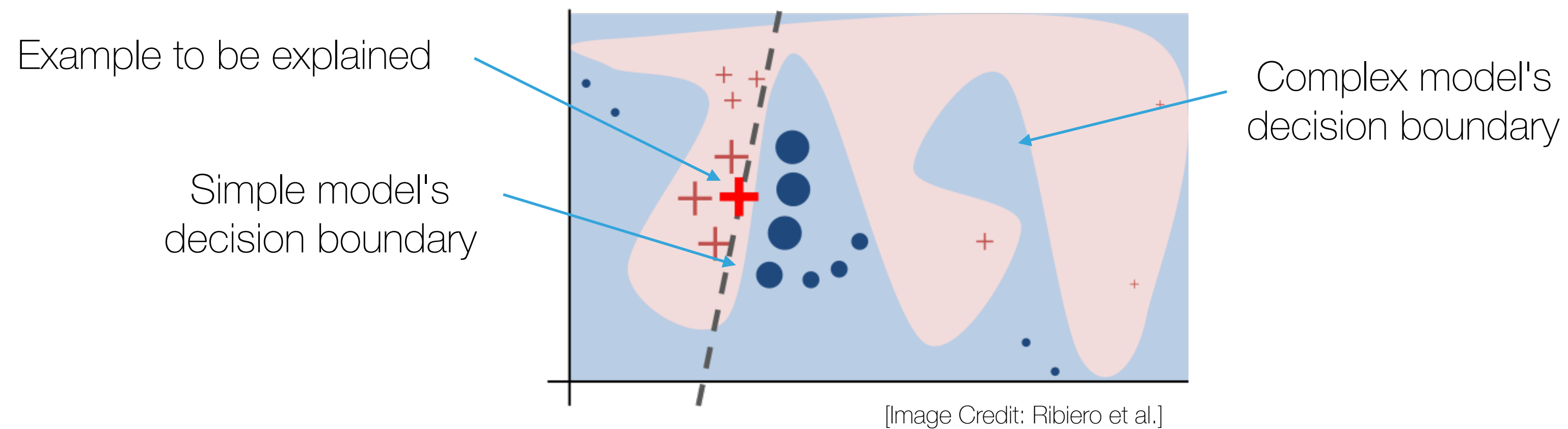
Simple model's
decision boundary



Examples to be explained

Interpretability through **Explanations**

- Perturbations [Ribeiro et al, '16; Lundberg & Lee, '17];



Explaining Structured Models