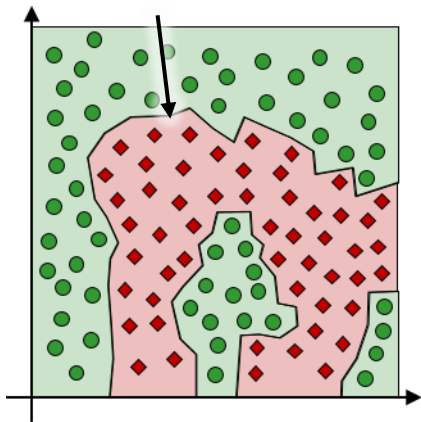


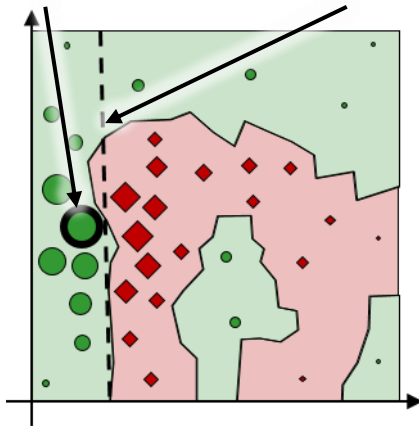
decision boundary of
black-box model

instance of interest

proxy model



original dataset



perturbed dataset



feature importance ranking