|  |  |
| --- | --- |
| Desiderata | How to test |
| Model-agnostic | - (by design) |
| Valid | Check that new prediction is of the target class |
| Robust | - (by design: deterministic) |
| Sparse | Count how many features are shown |
| Similar | Calculate the distance to the data point |
| Realistic | Calculate the distance to training data point |
| Informative | Show that the proposed feature changes are in line with the SHAP values |

Smart way to generate sparser explanations;

* Only show feature changes that are in line with the SHAP values.

Qualitative evaluation