index

For data $X_{n\times p}$, the projection basis A_{p*d} gives the projected data, $y = XA$. Projection pursuit finds the projection direction A that maximises the index function f:

$$\max f(XA) \quad A'A = 1.$$

The following sections summarise the index functions used in the simulation:

hole index

- smooth
- formula:

$$I_{holes}(A) = \frac{1 - 1/n \sum_{i=1}^{n} \exp(-1/2 y_i y_i')}{1 - \exp(-p/2)}$$

- reference:
  - Cook and Swayne (2007) Interactive and Dynamic Graphics for Data Analysis page 30: link
  - code from the tourr package

Stringy

- non-smooth
- formula:

$$I_{stringy}(A) = \frac{\text{number of vertices with 2 edges}}{\text{number of total vertices with more than one edge}}$$

  (probably need some graph theory notation to write it mathematically)

- reference:

  - Wilkinson et al (2005) Graph-Theoretic Scagnostics: page 160: link
  - code from the cassowaryr page

spline

- smooth
- formula:
  - $\hat{y}$ is derived from fitting a penalised spline model of y ~ x,
  - $\hat{y}'$ is derived from fitting a penalised spline model of x ~ y,
  - the variance of the residuals is compared to the variance of the original data

$$I_{spline}(A) = \max\left(1 - \frac{var(y - \hat{y})}{var(x)}, 1 - \frac{var(y - \hat{y}')}{var(y)}\right)$$

- reference:
  - Ursula and Di (2020) Using tours to visually investigate properties of new projection pursuit indexes with application to problems in physics, page 1176: link
  - code from the cassowaryr page