## Mutual information measurement

After describing the non-parametric formulation of the quadratic Renyi entropy combined with Parzen window estimation, we introduce the Cauchy-Schwartz mutual information, which can be directly numerically computed from the data. We use P to denote a probability and p for a probability density. Renyi entropy is a parametric family of entropy measures including Shannon entropy as a special case (i.e., order $\alpha \to 1$). The Renyi entropies of order $\alpha$ ($> 0, \neq 1$) of a discrete variable C and a continuous target Y are defined as:

$$H_{R_\alpha}(C) = \frac{1}{1-\alpha} log \sum_c p(c)^\alpha; \qquad H_{R_\alpha}(Y) = \frac{1}{1-\alpha} log \int_y p(y)^\alpha dy \tag{1}$$

To estimate in a non-parametric fashion the entropy measure of continuous variables, we used the quadratic order $\alpha = 2$ and the Parzen window method that introduces a kernel function centered at each sample. Therefore, the density of Y is estimated as a sum of spherical Gaussians:

$$\begin{gathered} p(y) = \frac{1}{N} \sum_{i=1}^{N} G(y - y_i, \sigma^2) \text{ with} \\ G(w, h^2) = (2\pi h^2)^{-\frac{1}{2}} \exp(-\frac{w^2}{2h^2}), \ h \text{ is the window width} \end{gathered} \tag{2}$$

$$\int_y G(y - u_1, \sigma_1^2) G(y - u_2, \sigma_2^2) dy = G(u_1 - u_2, \sigma_1^2 + \sigma_2^2) \tag{3}$$

Using the property of the convolution of two gaussian kernels (see Equation (3)), the quadratic Renyi entropy can be estimated as a sum of local interactions defined by the gaussian kernels centered at each pair of samples:

$$
\begin{aligned}
H_{R_2}(Y) &= -log \int_y p(y)^2 dy \\
&= -log \frac{1}{N^2} \int_y (\sum_{i=1}^{N} \sum_{j=1}^{N} G(y-y_i, \sigma^2) G(y-y_j, \sigma^2)) dy \\
&= -log \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} G(y_i - y_j, 2\sigma^2)
\end{aligned}
\tag{4}
$$

Finally, we expressed the mutual information between a discrete variable C and a continuous target Y in terms of quadratic Renyi entropies with the Cauchy-Schwartz divergence between the joint distribution $p(c, y)$ and the product of the marginal distributions $p(c)$ and $p(y)$ (see Equation (5)). The Cauchy-Schwartz mutual information $I_{CS}(C, Y) \geq 0$ with equality if and only if $p(c, y) = p(c)p(y)$ almost everywhere, that is if C and Y are independent. We normalized the mutual information score by dividing it with $0.5(I_{CS}(C, C) * I_{CS}(Y, Y))$. Hence, this normalized expression measures the average fraction of reduction in the uncertainty of one variable given the knowledge of the other variable. It ensures the property of symmetry and that the values fall within the unit range [0,1]

$$
\begin{aligned}
I_{CS}(C,Y) &= -log \frac{\sum_c \int_y p(c,y) * p(c)p(y) dy}{\sqrt{(\sum_c \int_y p(c,y)^2 dy)(\sum_c \int_y p(c)^2 p(y)^2 dy)}} \\
&= H_{R_2}(p(c,y) * p(c)p(y)) - \frac{1}{2} H_{R_2}(p(c,y)) - \frac{1}{2} H_{R_2}(p(c)p(y))
\end{aligned}
\tag{5}
$$