Suppose that you are trying to write a model that matches textual descriptions to images: it takes as input a pair \((x_{\text{im}}, x_{\text{text}})\) and outputs \(y = 1\) if the image is relevant to the textual description, and \(y = -1\) if not.

We want a nice\(^1\) kernel \(\kappa\) for data of the form \((x_{\text{im}}, x_{\text{text}})\), so that we can use kernel logistic regression to solve this problem.

Assume we know a nice kernel for image data, \(\kappa_{\text{im}}\), corresponding to a \(D_{\text{im}}\)-dimensional informative nonlinear feature map \(\phi_{\text{im}}\) for images, and that we know a nice kernel for text data, \(\kappa_{\text{text}}\), corresponding to a \(D_{\text{text}}\)-dimensional informative nonlinear feature map \(\phi_{\text{text}}\) for textual data.

I claim that there are at least two natural choices for a nice kernel for the mixed domain data:

\[
\begin{align*}
1) \quad & \kappa_{\text{mixed}}((x_{\text{im}}, x_{\text{text}}), (y_{\text{im}}, y_{\text{text}})) = \kappa_{\text{im}}(x_{\text{im}}, y_{\text{im}}) + \kappa_{\text{text}}(x_{\text{text}}, y_{\text{text}}) \\
2) \quad & \kappa_{\text{mixed}}((x_{\text{im}}, x_{\text{text}}), (y_{\text{im}}, y_{\text{text}})) = \kappa_{\text{im}}(x_{\text{im}}, y_{\text{im}}) \cdot \kappa_{\text{text}}(x_{\text{text}}, y_{\text{text}})
\end{align*}
\]

Write feature maps for these two mixed domain kernels in terms of the image and text feature maps, \(\phi_{\text{im}}\) and \(\phi_{\text{text}}\).

\(^1\)‘Nice’ means the kernel can be evaluated in time linear in the size of the input, as we saw for the gaussian and polynomial kernels: even though the feature maps for these two kernels have \(\omega(d)\) features, they can be computed in \(O(d)\) time, which is just the time it takes to read the data.