WEEKLY PARTICIPATION 7

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

Assume you know a nice kernel for image data, κ_{im}, corresponding to a D_{im}-dimensional informative nonlinear feature map ϕ_{im} for images, and that you know a nice kernel for text data, κ_{text}, corresponding to a D_{text}-dimensional informative nonlinear feature map ϕ_{text} for textual data.

You want a nice[1] kernel κ_{mixed} for data of the form $(\mathbf{x}_{\text{im}}, \mathbf{x}_{\text{text}})$, so that we can use kernel logistic regression to solve this problem, to get a model of the form

$$f(\boldsymbol{x}_{\text{im,new}}, \boldsymbol{x}_{\text{text,new}}) = \sum_{i=1}^{n_{\text{train}}} \alpha_i \kappa_{\text{mixed}}((\boldsymbol{x}_{\text{im,new}}, \boldsymbol{x}_{\text{text,new}}), (\boldsymbol{x}_{\text{im},i}, \boldsymbol{x}_{\text{text},i}))$$

Remember that these kernels correspond to some choice of feature maps ϕ_{im} and ϕ_{text}. You could form a new feature map by concatenating these feature maps, then form a new kernel function κ_{mixed} using this concatenated feature map. However, also remember that one of the goals of kernel learning is to avoid working with or even really thinking about the underlying feature maps. Hence we prefer to construct kernels from other kernels, eschewing the underlying feature maps.

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

(1) $\kappa_{\text{sum}}((\mathbf{x}_{\text{im}}, \mathbf{x}_{\text{text}}), (\mathbf{y}_{\text{im}}, \mathbf{y}_{\text{text}})) = \kappa_{\text{im}}(\mathbf{x}_{\text{im}}, \mathbf{y}_{\text{im}}) + \kappa_{\text{text}}(\mathbf{x}_{\text{text}}, \mathbf{y}_{\text{text}})$

(2) $\kappa_{\text{prod}}((\mathbf{x}_{\text{im}}, \mathbf{x}_{\text{text}}), (\mathbf{y}_{\text{im}}, \mathbf{y}_{\text{text}})) = \kappa_{\text{im}}(\mathbf{x}_{\text{im}}, \mathbf{y}_{\text{im}}) \cdot \kappa_{\text{text}}(\mathbf{x}_{\text{text}}, \mathbf{y}_{\text{text}})$

Demonstrate that these are valid kernels for the mixed domain:

- Write feature maps ϕ_{sum} and ϕ_{prod} for these two mixed domain kernels in terms of the image and text feature maps, ϕ_{im} and ϕ_{text}.

  You may find the column-wise Kronecker product (aka the Khatri-Rao) product useful for concisely writing one of the feature maps.

- Explain why the existence of these feature maps implies that κ_{sum} and κ_{prod} are valid kernels on the mixed domain.

- If $\phi_{\text{im}} : \mathbb{R}^d \to \mathbb{R}^{D_{\text{im}}}$ and $\phi_{\text{text}} : \mathbb{R}^d \to \mathbb{R}^{D_{\text{text}}}$, what are the dimensionalities D_{sum} and D_{prod} of, respectively, ϕ_{sum} and ϕ_{prod}?

Explain which of these kernels you think is more powerful, and why.

[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, the kernel function can be computed in $O(d)$ time, which is just the time it takes to read the two points being compared.