WEEKLY PARTICIPATION 9

Suppose that you are trying to write a 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.

We want a nice[1] kernel κ 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.

Assume we 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 we know a nice kernel for text data, κ_{text}, corresponding to a D_{text}-dimensional informative nonlinear feature map ϕ_{text} for textual data.

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

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

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

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