Multiple kernel multiple output learning

Abstract

1 Method

We use $\boldsymbol{\psi}(\mathbf{x}, \mathbf{y})$ to denote the joint feature map, which is composed by an input feature map $\boldsymbol{\phi}(\mathbf{x})$ and an output feature map $\boldsymbol{\varphi}(\mathbf{y})$ defined as

$$\boldsymbol{\psi}(\mathbf{x}, \mathbf{y}) = \boldsymbol{\phi}(\mathbf{x}) \otimes \boldsymbol{\varphi}(\mathbf{y}),$$

where $\otimes$ is the Kroecker product. In particular, the input feature is a collection of n feature maps defined as

$$\boldsymbol{\phi}(\mathbf{x}) = (\boldsymbol{\phi}_k(\mathbf{x}))_{k \in \{1, \cdots, n\}},$$

where $\boldsymbol{\phi}_k(\mathbf{x})$ is the kth input feature map, e.g., the bag-of-words feature of an example $\mathbf{x}$. The output feature map $\boldsymbol{\varphi}(\mathbf{y})$ is a collection of feature maps defined on the edges of a Markov network

$$\boldsymbol{\varphi}(\mathbf{y}) = (\boldsymbol{\varphi}_e(\mathbf{y}_e))_{e \in E},$$

where edge specific feature map $\boldsymbol{\varphi}_e(\mathbf{y}_e)$ is defined on edge $e = (p, q)$ as

$$\boldsymbol{\varphi}_e(\mathbf{y}_e) = (u_{e,nn}(\mathbf{y}_e), u_{e,np}(\mathbf{y}_e), u_{e,pn}(\mathbf{y}_e), u_{e,pp}(\mathbf{y}_e))^\intercal \otimes \boldsymbol{\theta}^{p,q}$$