Revisit literature

Thursday, February 13, 2025 11:45 AM

Multi-task Gaussian Process Prediction

Edwin V. Bonilla, Kian Ming A. Chai, Christopher K. I. Williams

However, as pointed out in [1] and supported empirically by [2], assuming relatedness in a set of tasks and simply learning them together can be detrimental.

We propose a model that attempts to learn inter-task dependencies based solely on the *task identities* and the observed data for each task.

This contrasts with approaches in [3, 4] where task-descriptor features t were used in a parametric covariance function over different tasks—such a function may be too constrained by both its parametric form and the task descriptors to model task similarities effectively. In [4, 3] this property also holds, but instead of specifying a general PSD matrix K^f, these authors set $K^f = k^f(t_l, t_k)$, where $k^f(\cdot, \cdot)$ is a covariance function over the task-descriptor features t.
Hence we propose a model that learns a "free-form" task-similarity matrix, which is used in conjunction with a parameterized covariance function over the input features x.

https://proceedings.mlr.press/v2/bonilla07a/bonilla07a.pdf

Evgeniou et al. [19] consider methods for inducing correlations between tasks based on a correlated prior over linear regression parameters. In fact this corresponds to a GP prior using the kernel $k(x, x') = x^T Ax'$ for some positive definite matrix A. In their experiments they use a restricted
form of K^f with $K^f = (1 - \lambda) + \lambda M\delta_{lk}$ (their eq. 25), i.e. a convex combination of a rank-1 $_{lk}$
matrix of ones and a multiple of the identity. Notice the similarity to the PPCA form of K^f given in section 3.

Finally, we make use of GP approximations and properties of our model in
order to scale our approach to large multi-task data sets