

Phylogenetic Regularised Regression Maths

Russell Dinnage

17/11/2020

Bayesian Regularised Regression on Phylogenies (BRRP)

A vector valued trait in species i , T_i can be modeled as:

$$T_i = \alpha + \sum_j^{n_{\text{edge}}} \beta_j \gamma_j(i)$$

where n_{edge} is the number of edges in the phylogeny and

$$\gamma_j(i) = \begin{cases} 0 & j \notin P_i \\ l_j & j \in P_i \end{cases}$$

where P_i is the set of all edges leading from the root to species i , and l_j is the branchlength associated with edge j . Then α is the trait value at the root (equivalent to the intercept in linear regression), and the β_j parameters are the rates of evolution in the trait along edge j (equivalent to regression coefficients). This is an over-parameterized linear regression. Over-parameterization is solved through regularisation (or weakly informative priors in the Bayesian formulation).