

We propose a new method for supervised learning. The hubNet procedure fits a hub-based graphical model to the predictors, to estimate the amount of “connection” that each predictor has with other predictors. This yields a set of predictor weights that are then used in a regularized regression such as the lasso or elastic net. The resulting procedure is easy to implement, can often yield higher or competitive prediction accuracy with fewer features than the lasso and can give insights into the underlying structure of the predictors.

HubNet can be generalized seamlessly to supervised problems such as regularized logistic regression (and other GLMs), Cox’s proportional hazards model, and nonlinear procedures such as random forests and boosting. We prove recovery results under a specialized model and illustrate the method on real and simulated data.