

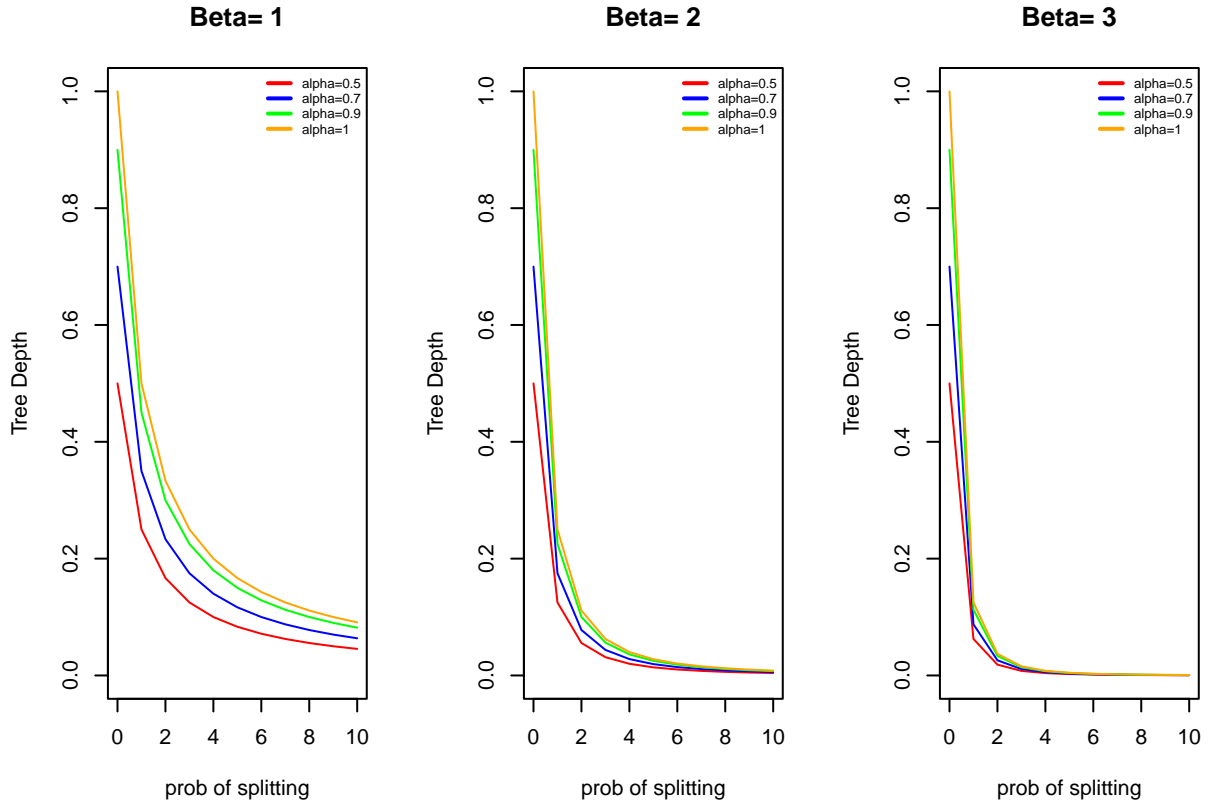
# BART

Bayesian Additive Regression Trees (BART) is a kind of tree-based model. In this model, we first grow many small subtrees and then sum the predictions of these small subtree models to get our final prediction. In Hill's article, she defines a subtree with a parameter pair  $(T, M)$ , where  $T$  denotes the tree and  $M$  denotes the parameters for the leaves of one tree.

To avoid overfitting, regularization priors are applied. Due to the independence assumption, we only need to define three priors for  $p(T)$ ,  $p(M|T)$ ,  $p(\sigma^2)$ .

First, We need to control the maximum depth of each subtree to avoid overfitting. Prior of  $p(T)$  controls the depth of each subtree. The probability that a node at depth  $d$  splits is

$$\alpha(1 + d)^{-\beta}, \alpha \in (0, 1), \beta \in [1, \infty)$$



The plots above show how different  $\alpha, \beta$  influence this probability. Generally speaking, when  $\alpha$  gets larger the probability that a node at depth  $d$  splits gets stronger. And when  $\beta$  gets larger this probability goes down faster as the tree goes deeper.

Second, we need to shrink the sum of the fit across trees toward zero. So we assume that the prior distribution for each means  $\mu$  follows a normal distribution.

$$\mu \sim N(0, \sigma_\mu^2), \sigma_\mu = 0.5/(k\sqrt{m})$$

and when  $k=2$ , the response variable has 95 probability lies between  $-0.5$  and  $0.5$ .

Third, the prior on  $\sigma^2$  is specified as an inverse  $\chi^2$  distribution. The inverse  $\chi^2$  distribution is the prior of the normal distribution when its variance is unknown.

In computation, instead of fitting the original response variable  $y$ , we fit the residual  $R$  ( $y$  - the predictions of the other subtrees). The subtrees are updated one by one several times. When updating a single subtree we first choose its  $T$  using the Metropolis-Hasting approach and then update its  $M$ . After several loops, we can get our final BART model.