

## TREE APPROXIMATORS OF A NORMAL

## Why

What is the optimal tree approximator of a multivariate normal density?

## Result

**Prop. 1.** Let  $g: \mathbb{R}^n \to \mathbb{R}$  be a normal density with mean  $\mu \in \mathbb{R}^d$  and covariance  $\Sigma \in \mathbb{S}^d_{++}$ . The normal density  $f_T^*: \mathbb{R}^d \to \mathbb{R}$  with mean  $\mu$  and precision matrix P defined by

• 
$$P_{11} = \Sigma_{11}^{-1} + \sum_{\mathbf{pa}} \sum_{j=1}^{2} \Sigma_{j1}^{2} \Sigma_{11}^{-2} \Sigma_{j|1}^{-1}$$

• for 
$$i = 2, ..., d$$
,  $P_{ii} = \sum_{i|\mathbf{pa}}^{-1} \sum_{j=1}^{1} \sum_{j=1}^{2} \sum_{i=1}^{2} \sum_{j=1}^{1} \sum_{j=1}^{2} \sum_{$ 

• 
$$i, j = 1, \dots d \text{ and } i = \mathsf{pa} j, \ P_{ij} = P_{ji} = -\Sigma_{ji} \Sigma_{jj}^{-1} \Sigma_{j|i}^{-1}$$

where pai is the parent of i in an optimal approximator tree T (i = 2, ..., n) is an optimal tree approximator of g.

*Proof.* Using Proposition 1 of Best Tree Density Approximators, express an optimal tree approximator of g by

$$(1/c) \exp \left( -\frac{1}{2} \left( \sum_{11}^{-1} \bar{x}_1^2 + \sum_{i \neq 1} \left( \bar{x}_i - \sum_{i, \mathbf{pa}} \sum_{i}^{-1} \sum_{\mathbf{pa}} \bar{x}_{i, \mathbf{pa}} \bar{x}_{i} \bar{x}_{\mathbf{pa}} \right)^2 \sum_{i \mid \mathbf{pa}}^{-1} \right) \right)$$

where 
$$\bar{x}_i = x_i - \mu_i$$
 and  $c = \sqrt{(2\pi)^d \sum_{11} \prod_{i \neq 1} \sum_{i \mid \mathbf{pa} i}}$ .

Second, express the quadratic in the exponential as

$$\Sigma_{11}^{-1} \bar{x}_1^2 + \sum_{i \neq 1} \left[ \Sigma_{i \mid \mathbf{pa} \, i}^{-1} \bar{x}_i^2 - 2 \Sigma_{i, \mathbf{pa} \, i} \Sigma_{\mathbf{pa} \, i, \mathbf{pa} \, i}^{-1} \Sigma_{i \mid \mathbf{pa} \, i}^{-1} \bar{x}_i \bar{x}_{\mathbf{pa} \, i} + \Sigma_{i, \mathbf{pa} \, i}^2 \Sigma_{\mathbf{pa} \, i, \mathbf{pa} \, i}^{-2} \Sigma_{i \mid \mathbf{pa} \, i}^{-1} \bar{x}_{\mathbf{pa} \, i}^2 \Sigma_{\mathbf{pa} \, i, \mathbf{pa} \, i}^{-1} \Sigma_{i \mid \mathbf{pa} \, i}^{-1} \bar{x}_{\mathbf{pa} \, i}^2 \Sigma_{\mathbf{pa} \, i, \mathbf{pa} \, i}^{-1} \Sigma_{i \mid \mathbf{pa} \, i}^{-1} \bar{x}_{\mathbf{pa} \, i}^2 \Sigma_{\mathbf{pa} \, i, \mathbf{pa} \, i}^{-1} \Sigma_{i \mid \mathbf{pa} \, i}^{-1} \Sigma_{\mathbf{pa} \, i, \mathbf{pa} \, i}^{-1} \Sigma_{i \mid \mathbf{pa} \, i}^{-1} \Sigma_{\mathbf{pa} \, i, \mathbf{pa} \, i}^{-1} \Sigma_{i \mid \mathbf{pa} \, i}^{-1} \Sigma_{i \mid \mathbf{pa} \, i}^{-1} \Sigma_{\mathbf{pa} \, i, \mathbf{pa} \, i}^{-1} \Sigma_{i \mid \mathbf{pa} \, i}^{-1} \Sigma_{$$

With P defined as earlier, we can express the above as  $\bar{x}^{\top}P\bar{x}$ .

Third, note that c is  $\sqrt{(2\pi)^d \det P^{-1}}$  since  $f_T^*$  is a density and so integrates to one.

Notice that  $f_T^*$  is a tree normal density.

## **Empirical Normal**

In particular, notice that we can approximate the empirical normal density of a dataset with a density that factors according to a tree.

