Tree Approximators Of A Normal

1 Why

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

2 Result

Proposition 1. *Let $g : \mathsf{R}^n \to \mathsf{R}$ be a normal density with mean $\mu \in \mathsf{R}^d$ and covariance $\Sigma \in \mathbf{S}^d_{++}$. The normal density $f^*_T : \mathsf{R}^d \to \mathsf{R}$ with mean μ and precision matrix P defined by*

- $P_{11} = \Sigma_{11}^{-1} + \sum_{\mathsf{pa}_j = 1} \Sigma_{j1}^2 \Sigma_{11}^{-2} \Sigma_{j|1}^{-1}$
- *for $i = 2, \dots, d$, $P_{ii} = \Sigma_{i|\mathsf{pa}_i}^{-1} + \sum_{\mathsf{pa}_j = i} \Sigma_{ji}^2 \Sigma_{ii}^{-2} \Sigma_{j|i}^{-1}$*
- *$i, j = 1, \dots d$ and $i = \mathsf{pa}_j$, $P_{ij} = P_{ji} = -\Sigma_{ji}\Sigma_{jj}^{-1}\Sigma_{j|i}^{-1}$*

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

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

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

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

Second, express the quadratic in the exponential as

$$\Sigma_{11}^{-1}\bar{x}_1^2 + \sum_{i \neq 1}\left[\Sigma_{i|\mathsf{pa}_i}^{-1}\bar{x}_i^2 - 2\Sigma_{i,\mathsf{pa}_i}\Sigma_{\mathsf{pa}_i,\mathsf{pa}_i}^{-1}\Sigma_{i|\mathsf{pa}_i}^{-1}\bar{x}_i\bar{x}_{\mathsf{pa}_i} + \Sigma_{i,\mathsf{pa}_i}^2\Sigma_{\mathsf{pa}_i,\mathsf{pa}_i}^{-2}\Sigma_{i|\mathsf{pa}_i}^{-1}\bar{x}_{\mathsf{pa}_i}^2\right]$$

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

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

$\square$

Notice that f^*_T is a tree normal density.

3 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.