Want to estimate

\[ S \equiv \{ x \in [0, 1]^d : f(x) > \gamma \} \]

of a multidimensional function

\[ f : [0, 1]^d \rightarrow [C_\ell, C_u] \]

from noisy observations

\[ y_i = f(x_i) + \epsilon_i \]

\[ \mathbb{E}[y_i] = f(x_i) \]

\[ y_i \text{ bounded} \]
Error metric

Excess risk:

\[ R(T) - R(S) = 2 \int_{\Delta(T,S)} |f(x) - \gamma| dP_x \]

\[ \Delta(\square, \bullet) = \square \]

true level set

estimate slightly off, but in important regions

estimate off by more area, but in less significant regions
Balanced vs. Unbalanced Trees

same number of leaves

\[
\sum_{L \in T_1} 1 = \sum_{L \in T_2} 1
\]

\[
\sum_{L \in T_1} \sqrt{\text{area}(L) \cdot \text{depth}(L)} \ll \sum_{L \in T_2} \sqrt{\text{area}(L) \cdot \text{depth}(L)}
\]
Extracting information from complicated signals

- True field
- True contour
- Thresholded denoised field
- Noisy observations
- Thresholded observations
- Partition-based estimate