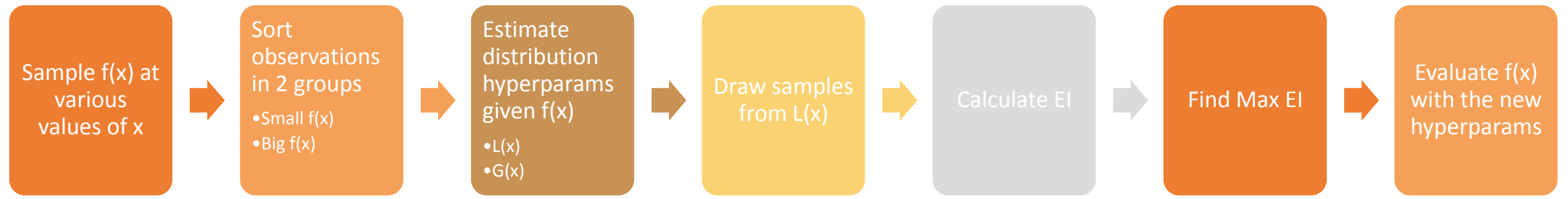


TPE Procedure

• TPE – Procedure

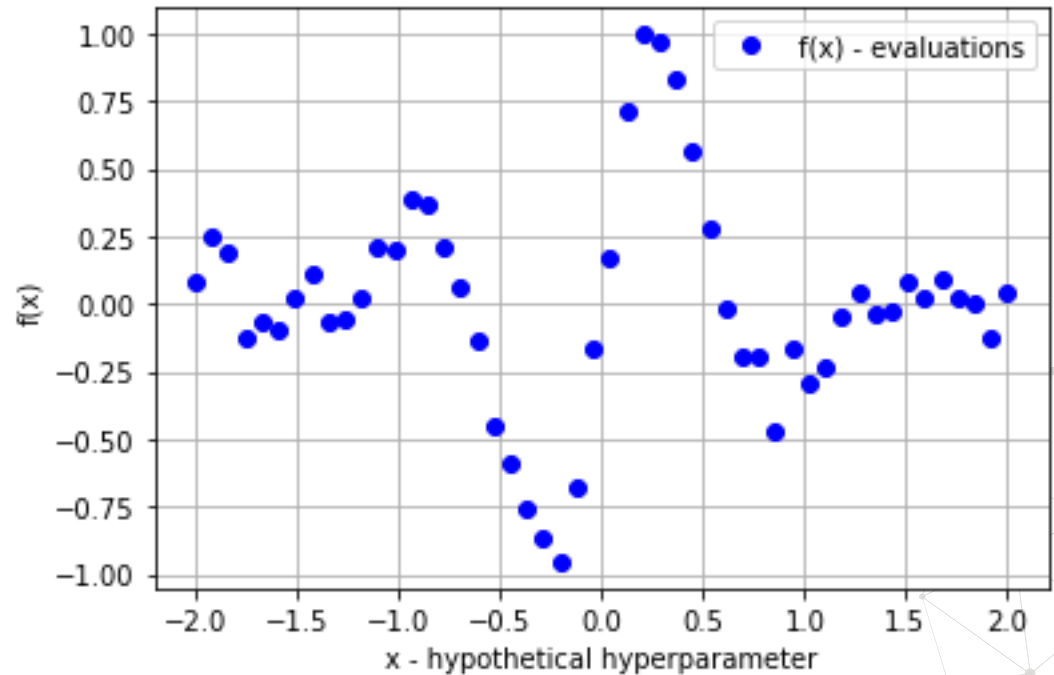


TPE – Step 1 – Sample $f(x)$

Evaluate $f(x)$ at different values of the hyperparameters

In this example, 1 hyperparameter

→ x

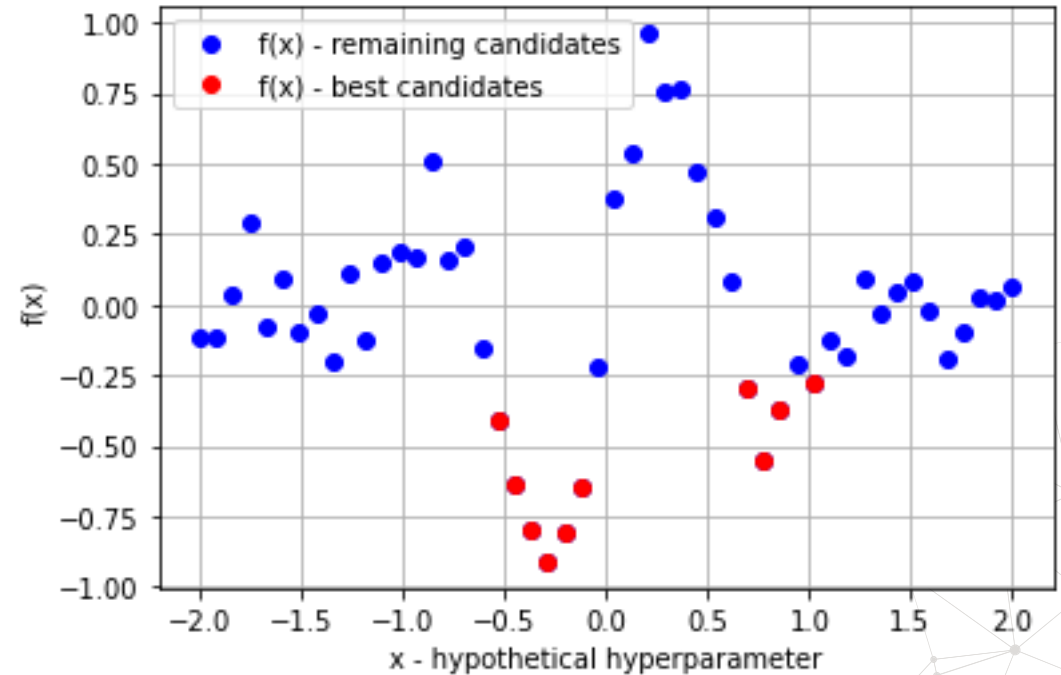


TPE – Step 2 – Divide observations

Separate observations in 2 groups:

First group: observations with best scores, smallest values of $f(x)$ (red)

Second group: the rest of the observations (blue)

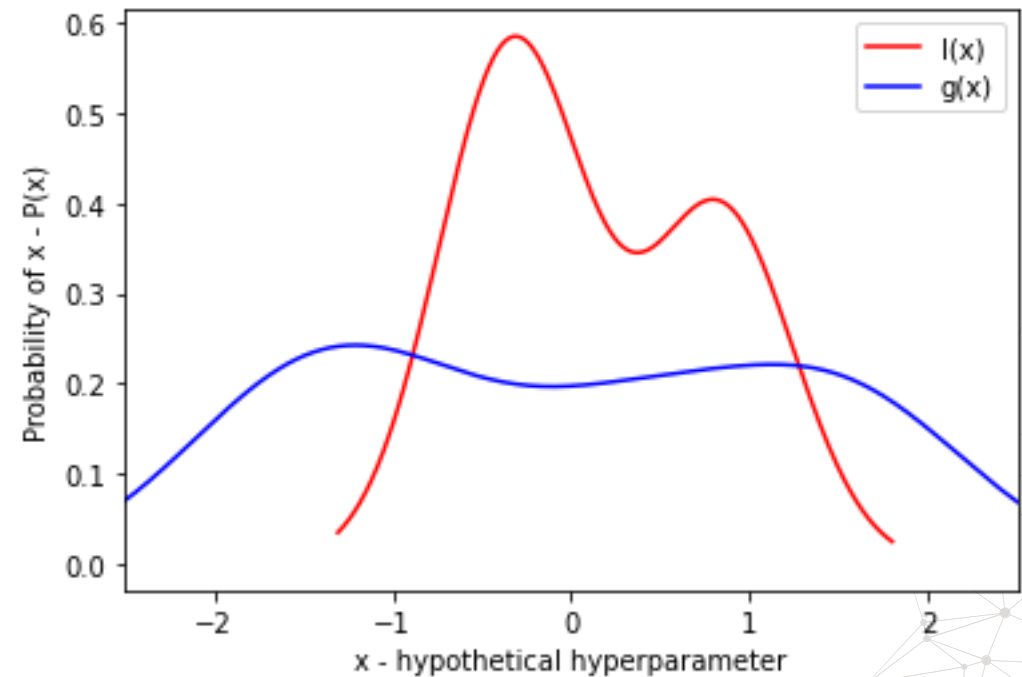


• TPE – Step 3 – Estimate distribution of x

Estimate the distributions of each hyperparameter and each group.

Distributions are estimated by **Parzen windows**, which is another name for kernel density estimation.

Thus, the name Parzen Estimators in TPE

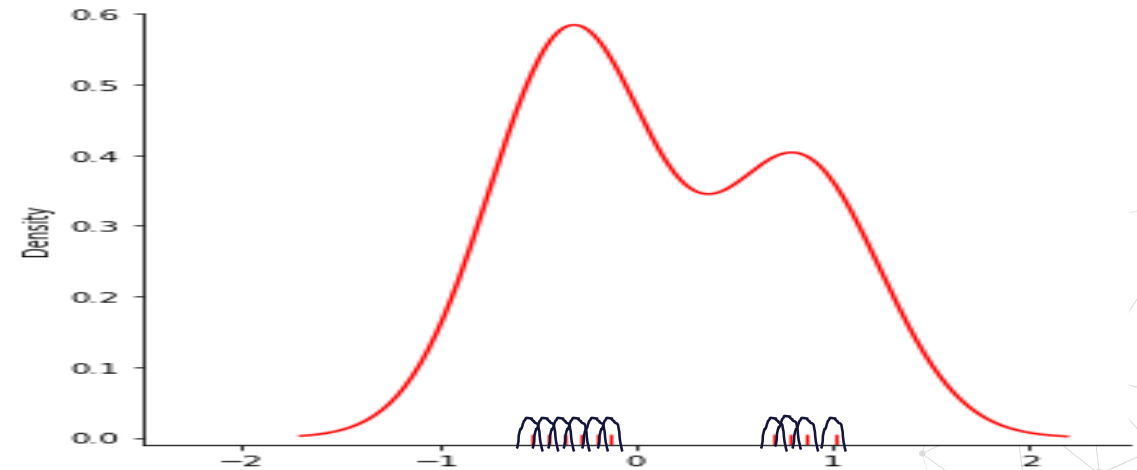


• TPE – Step 3 – Estimate distribution of x

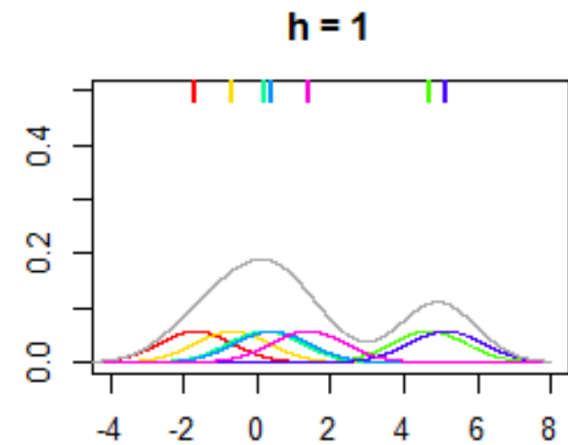
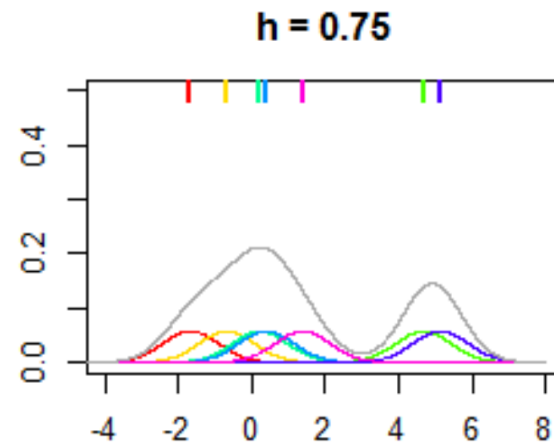
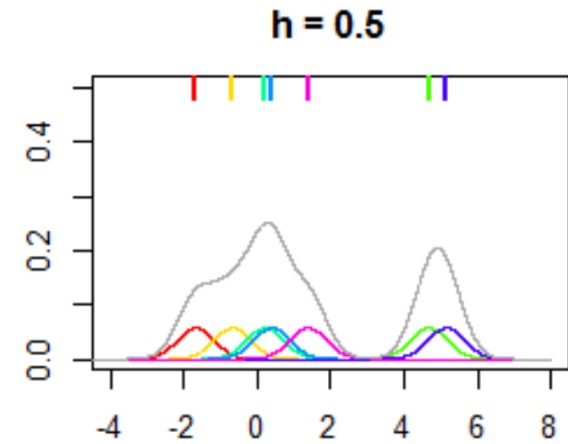
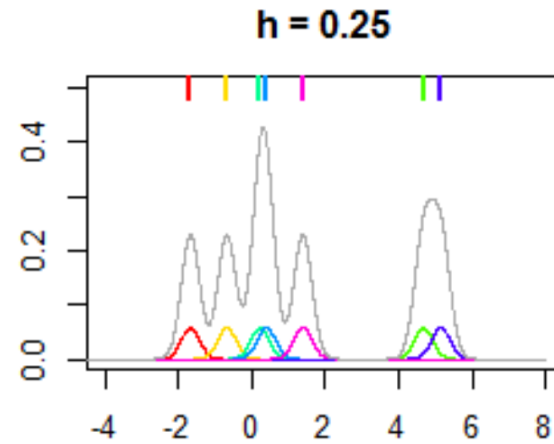
Estimate the distributions of each group.

Distributions are estimated by Parzen windows, which is another name for kernel estimation.

Thus, the name Parzen Estimators in TPE



Parzen Windows



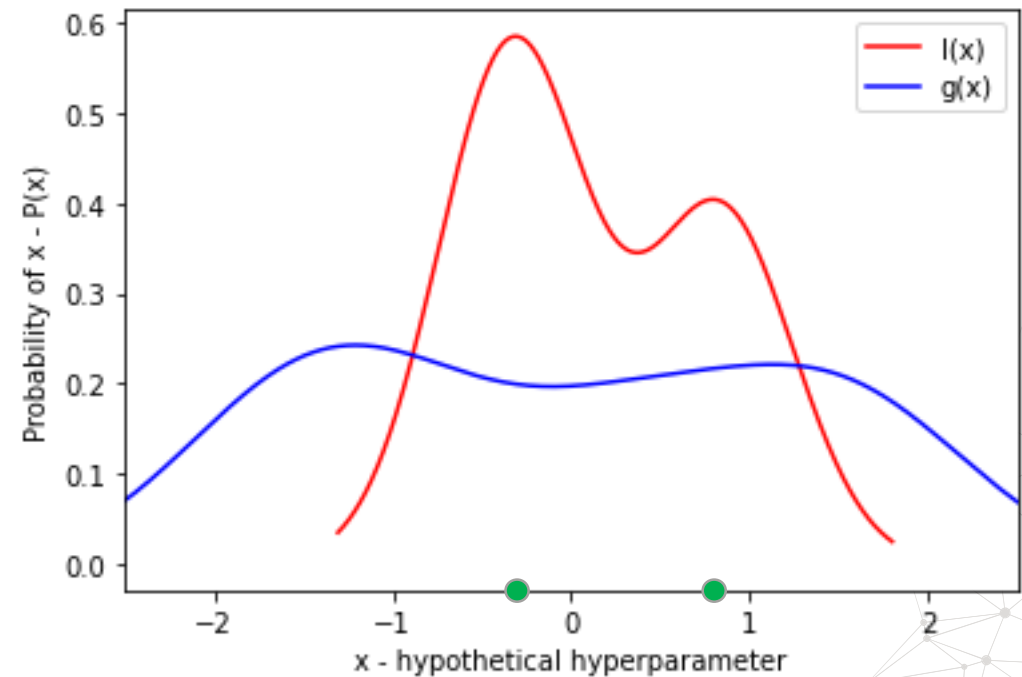
<https://stats.stackexchange.com/questions/244012/can-you-explain-parzen-window-kernel-density-estimation-in-laymans-terms>

TPE – Step 4 – Estimate EI

Draw samples from $L(x)$

Determine EI

Find maximum EI

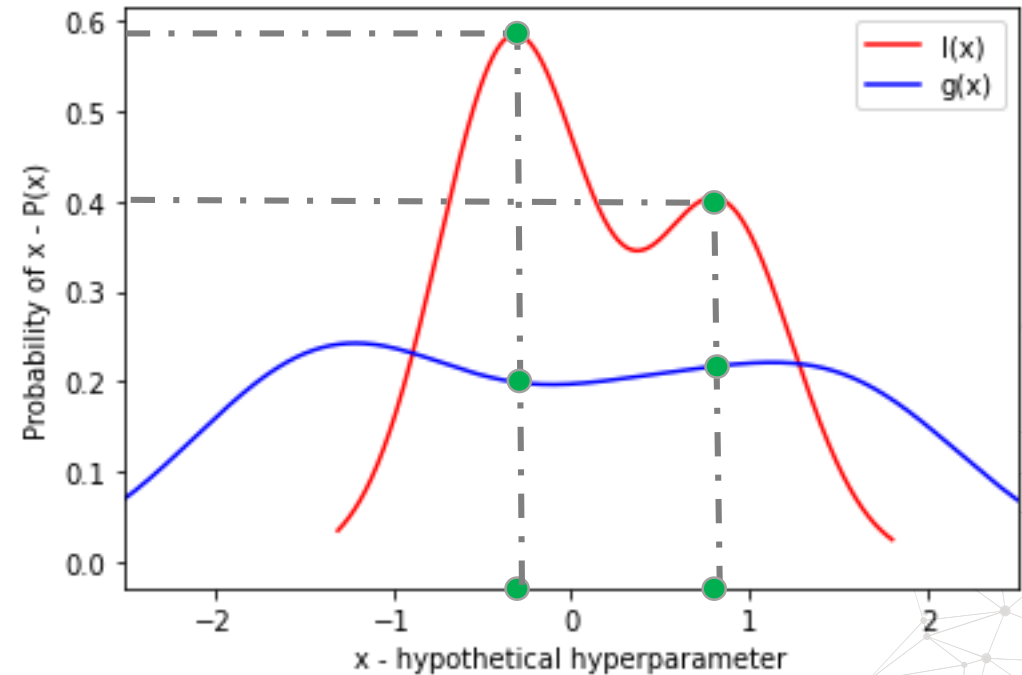


TPE – Step 4 – Estimate EI

Draw samples from $l(x)$

Determine EI $\propto \left(\gamma + \frac{g(x)}{\ell(x)} (1 - \gamma) \right)^{-1}$

Find maximum EI

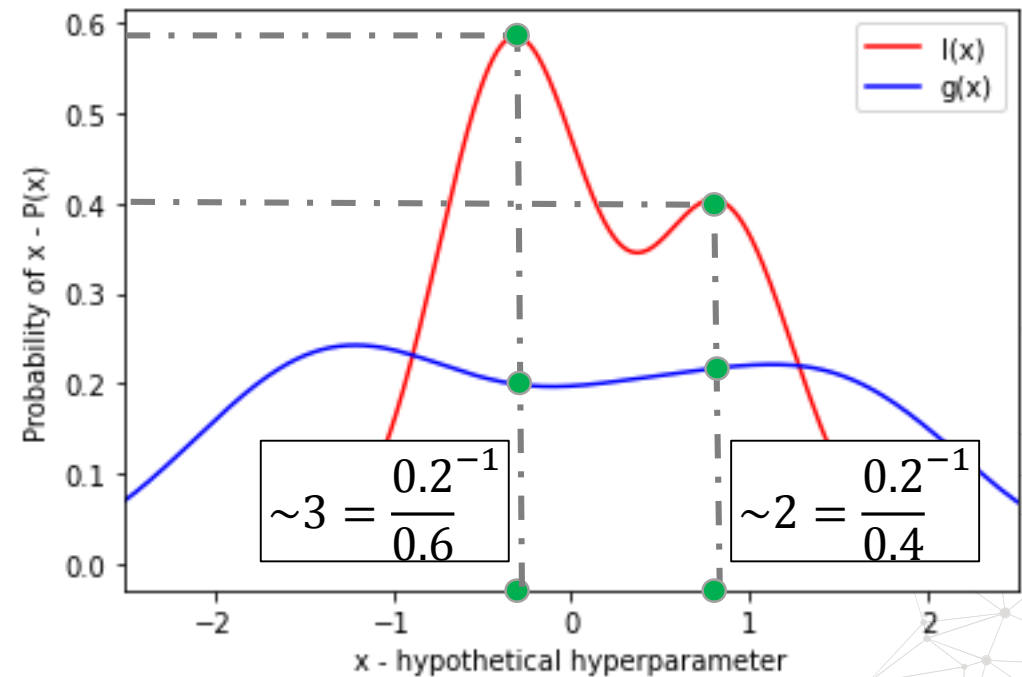


TPE – Step 4 – Estimate EI

Draw samples from $l(x)$

Determine EI $\propto \left(\gamma + \frac{g(x)}{\ell(x)}(1 - \gamma) \right)^{-1}$

Find maximum EI



To maximize EI we would like points x (Hyperparams) with high probability under $l(x)$ and low probability under $g(x)$

THANK YOU

www.trainindata.com