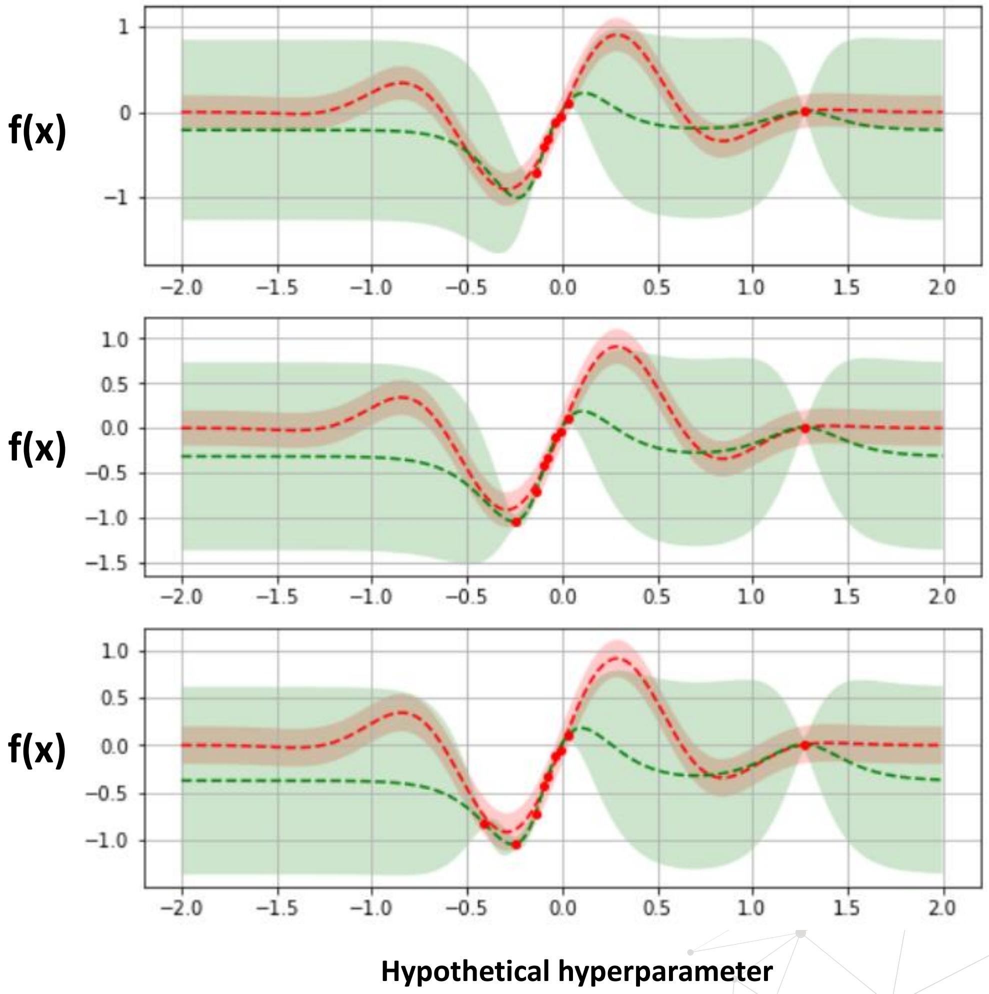
Train In Data

Sequential Model-Based Algorithm Configuration

SMAC

Bayesian Hyperparameter Optimization

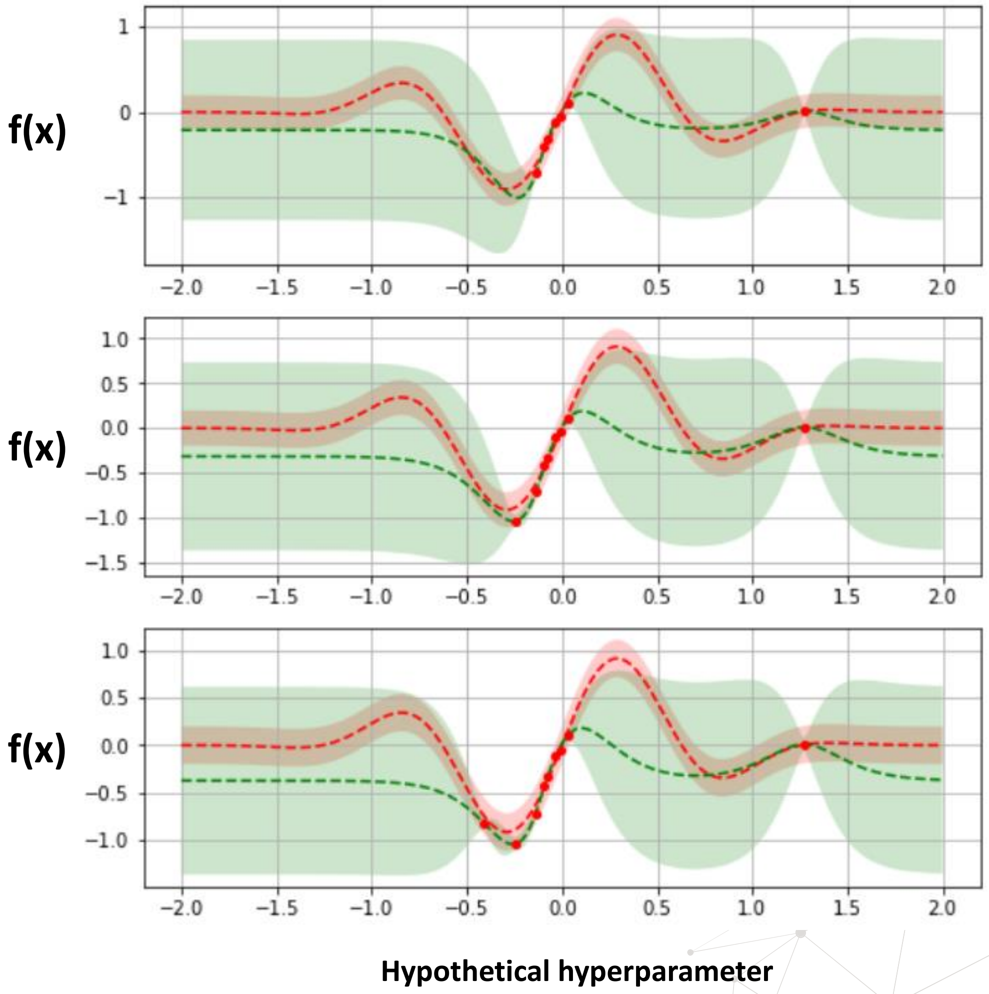
- *Red: f(x) – the hyperparameter response surface*
- *Green: our estimation of f(x)*
  - *In Bayes terms P(y|x) with y model score (f(x)) and x hyperparameters*



Bayesian Hyperparameter Optimization

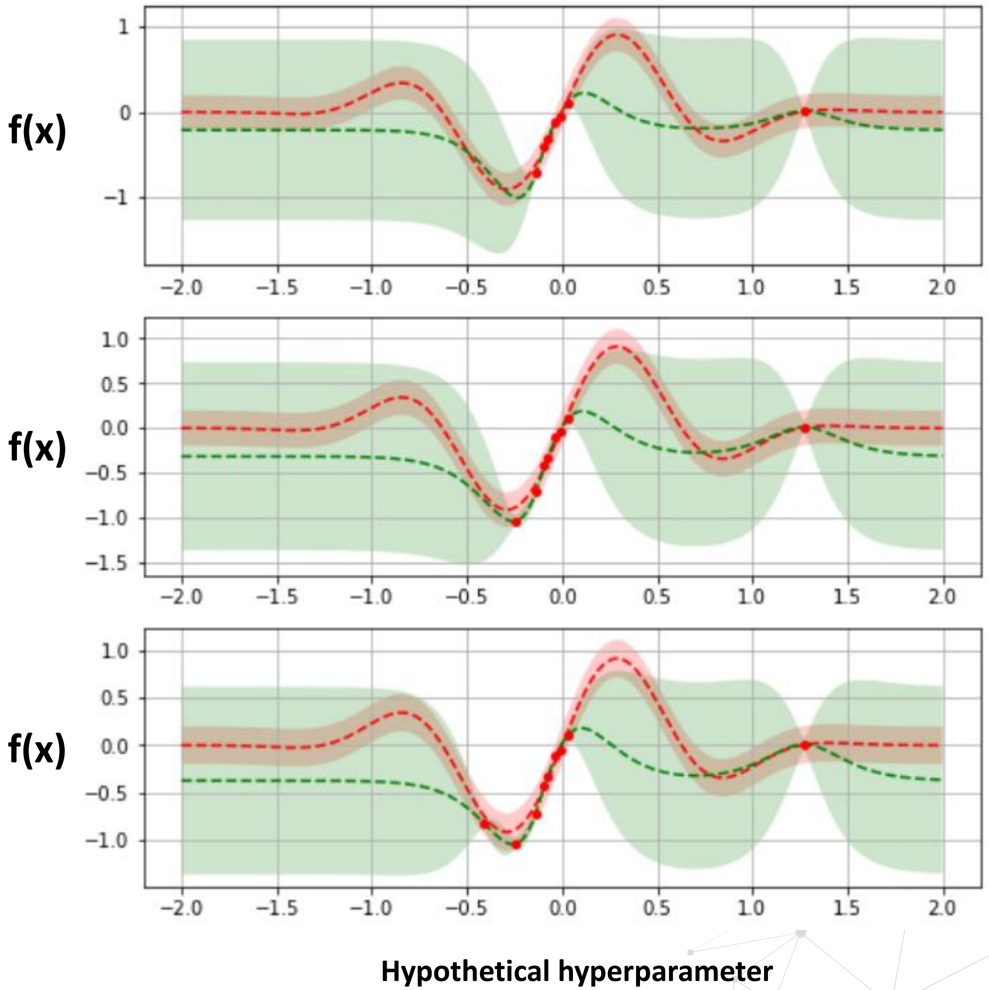
- *Red: f(x) – the hyperparameter response surface*

- *Green: our estimation of f(x)*
  - *In Bayes terms P(y|x) with y model score (f(x)) and x hyperparameters*
  - *Gaussian Process*



SMAC

- Red: f(x) – the hyperparameter response surface
- Green: our estimation of f(x)
  - In Bayes terms P(y|x) with y model score (f(x)) and x hyperparameters
  - Gaussian Process
  - ***Random Forests***
  - (Gradient boosted trees)



SMAC

- *Approximating f(x) is a regression problem*
- *Hyperparameters are the input "predictive" variables*
- *f(x) is our response variable or target*

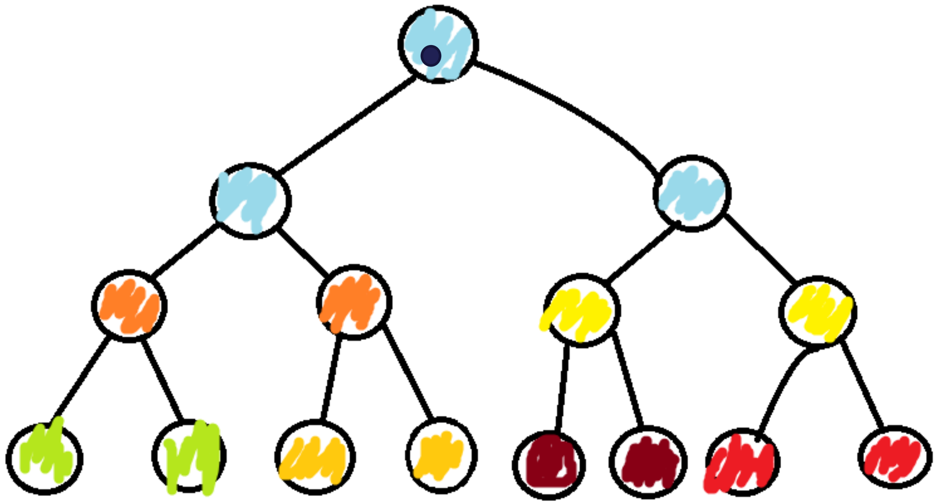
Hyperparam 1	Hyperparam 2	Hyperparam 3	Hyperparam 4	f(x) – model performance
0.1	1	512	32	0.65
0.01	2	48	16	0.78
0.005	3	12	8	0.89
0.03	2	250	1	0.66

SMAC

- *Approximating f(x) is a regression problem*
- *Hyperparameters are the input "predictive" variables*
- *f(x) is our response variable or target*
- *Predict f(x) for unseen values of the hyperparameters*
- *Decide where to sample next ➔ Acquisition functions*

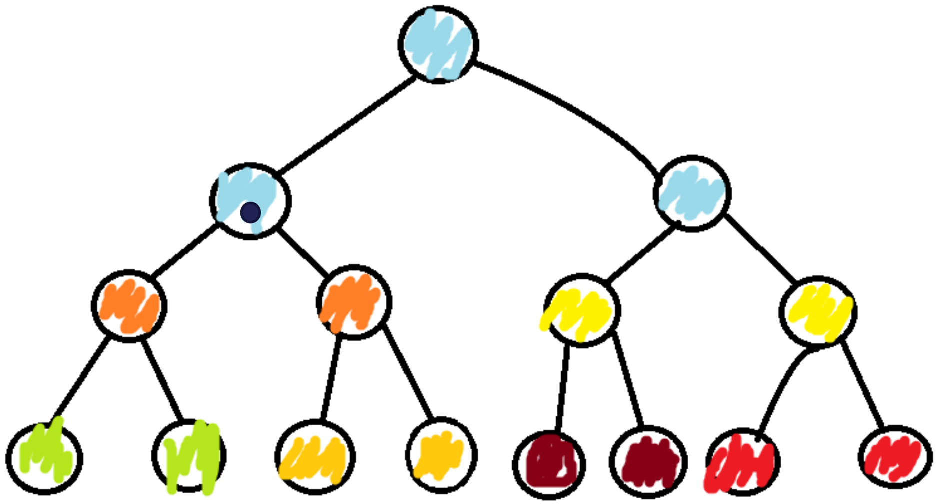
Hyperparam 1	Hyperparam 2	Hyperparam 3	Hyperparam 4	f(x) – model performance
0.1	1	512	32	0.65
0.01	2	48	16	0.78
0.005	3	12	8	0.89
0.03	2	250	1	0.66

Decision Tree – mean and error



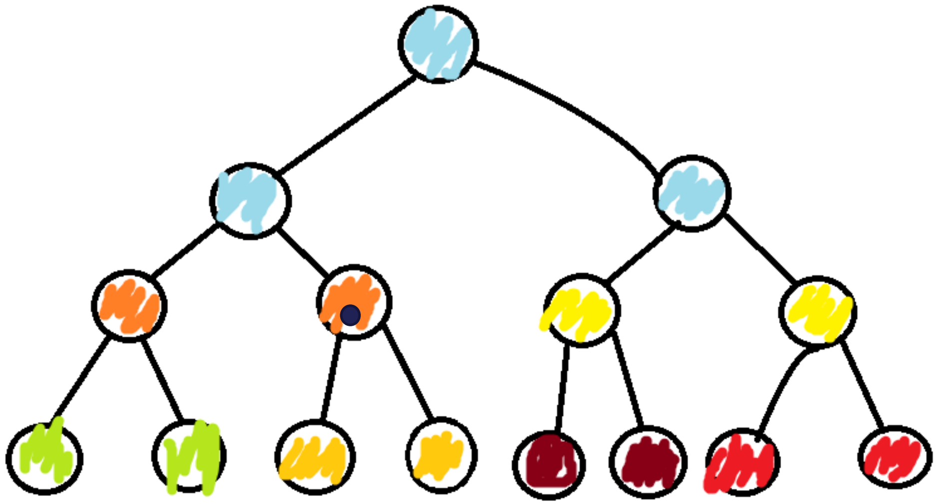
- With a trained decision tree, every new observation will be allocated to an end node

Decision Tree – mean and error



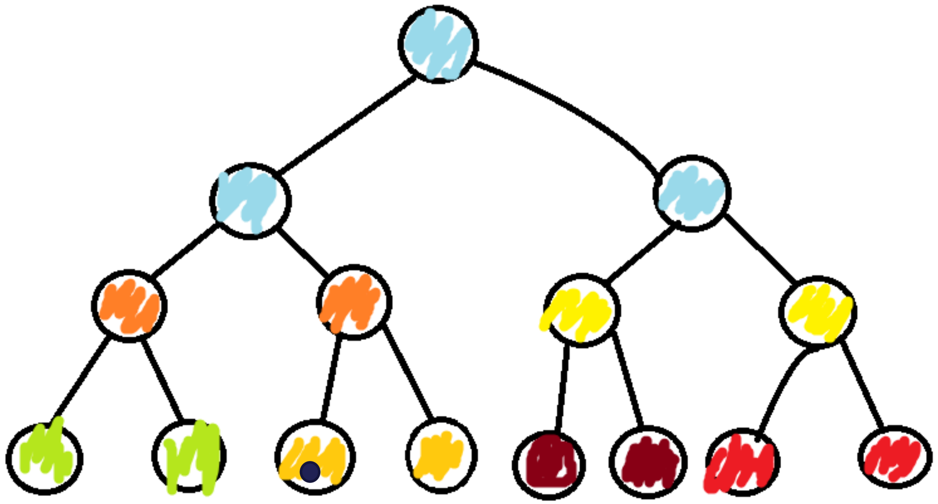
- With a trained decision tree, every new observation will be allocated to an end node

Decision Tree – mean and error



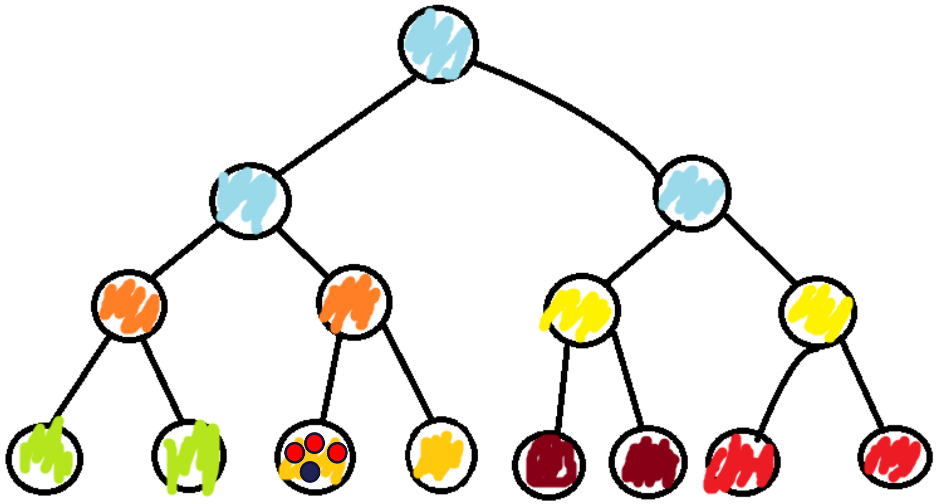
- With a trained decision tree, every new observation will be allocated to an end node

Decision Tree – mean and error



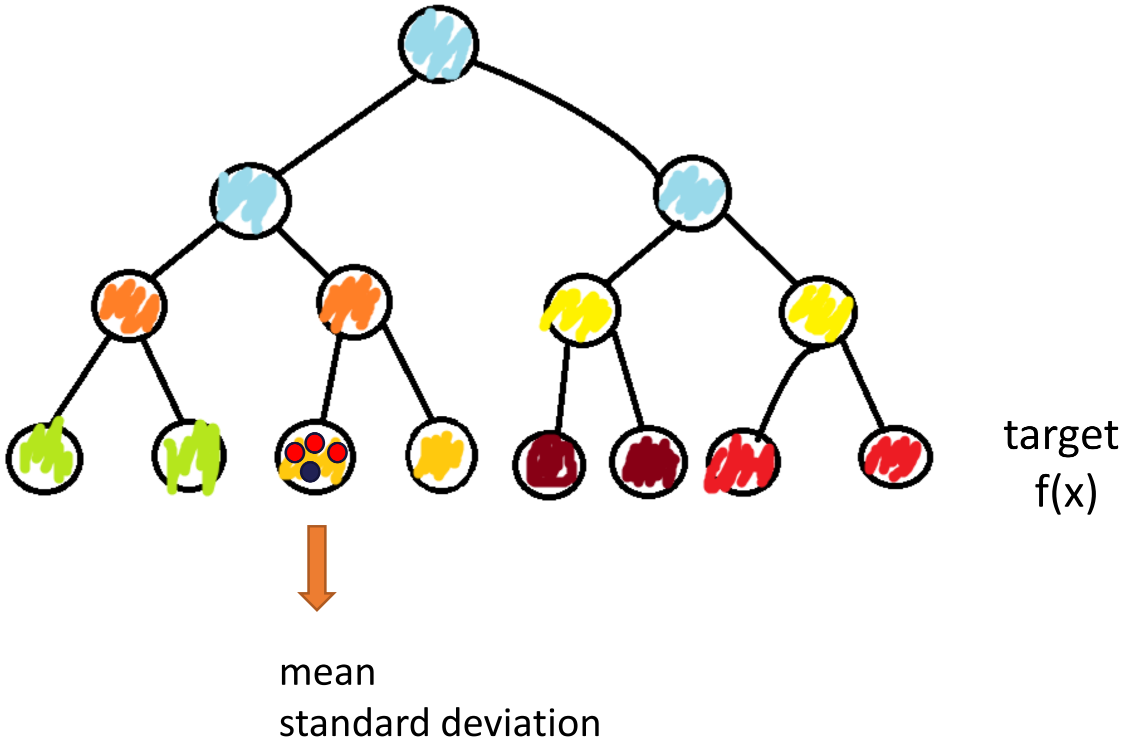
- With a trained decision tree, every new observation will be allocated to an end node

Decision Tree – mean and error



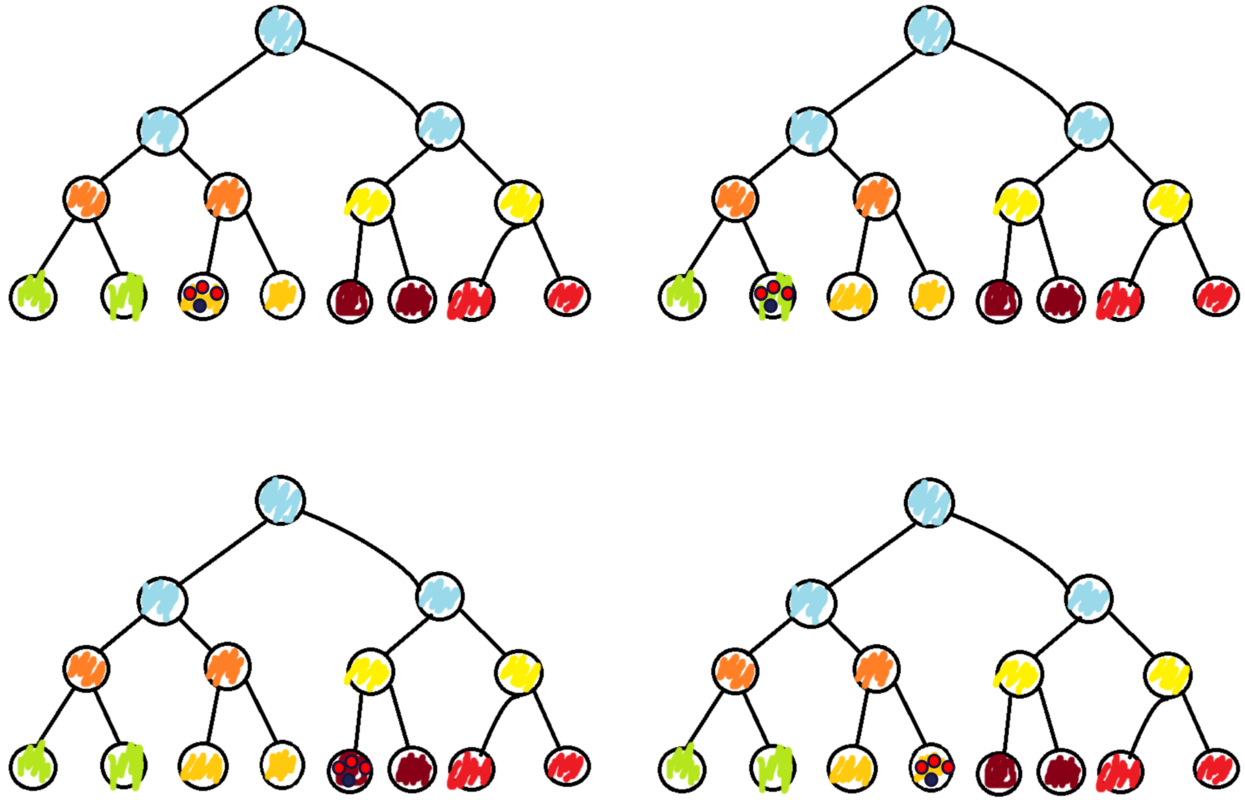
- With a trained decision tree, every new observation will be allocated to an end node
- In the trained tree, each end node contains a bunch of training observations

Decision Tree – mean and error



- With a trained decision tree, every new observation will be allocated to an end node
- In the trained tree, each end node contains a bunch of training observations
  - It is possible to derive a mean and standard deviation for f(x)

Random Forests – mean and error



- With a trained decision tree, every new observation will be allocated to an end node
- In the trained tree, each end node contains a bunch of training observations
    - It is possible to derive a mean and standard deviation for f(x)

Acquisition Functions – Scikit-Optimize

- Probability of Improvement (PI)
- Expected Improvement (EI)
- Upper (or Lower) confidence bound (UCB or LCB)

Acquisition Functions – SMAC

- Expected Improvement (EI) of the log transformed cost

$$EI(\boldsymbol{\theta}) := E[I_{\exp}(\boldsymbol{\theta})] = f_{min}\Phi(v) - e^{\frac{1}{2}\sigma_{\boldsymbol{\theta}}^2 + \mu_{\theta}} \cdot \Phi(v - \sigma_{\boldsymbol{\theta}}),$$

$$\Downarrow$$

$$v := \frac{\ln(f_{min}) - \mu_{\theta}}{\sigma_{\theta}},$$

THANK YOU

www.trainindata.com