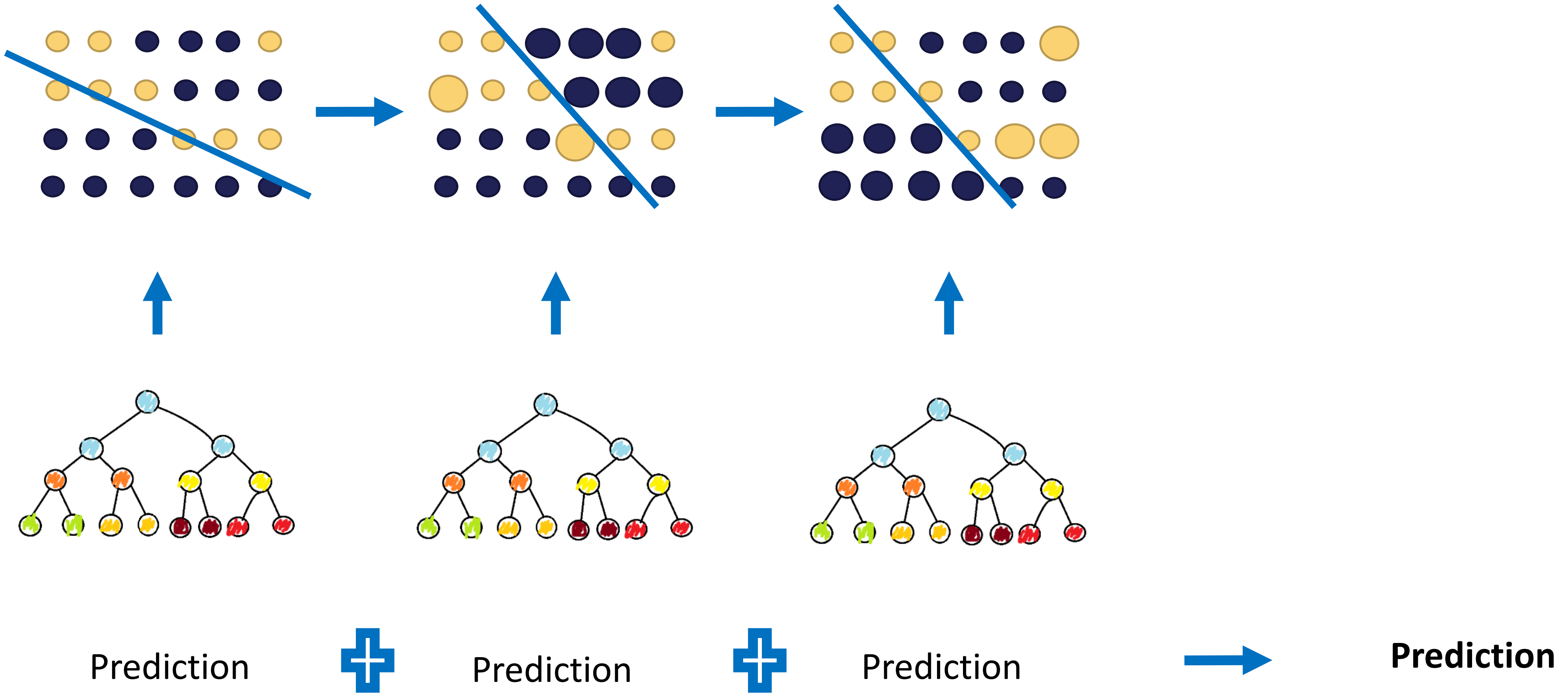
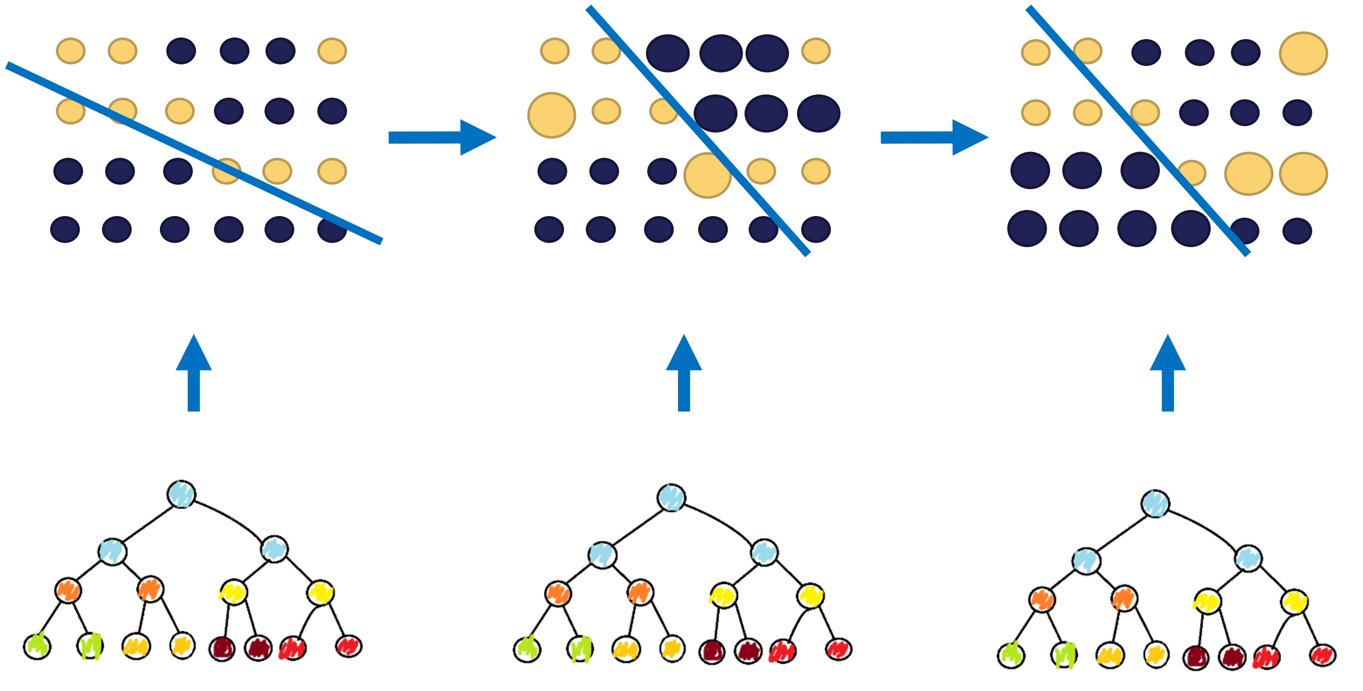
Explaining GBMs locally

GBMs



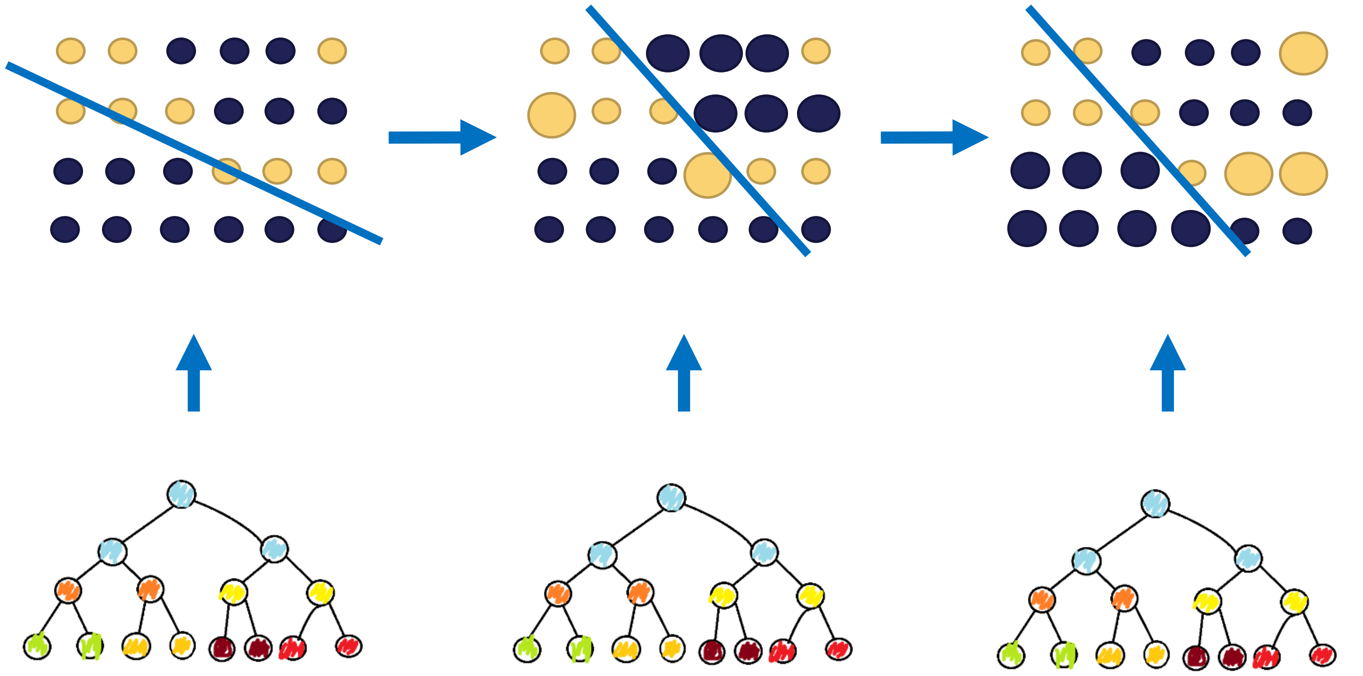
GBMs



$$f(x) = \sum_{m=1}^{M} \beta_m b(x; \gamma_m),$$

Prediction + Prediction + Prediction → **Prediction**

GBMs



Prediction + Prediction + Prediction → **Prediction**

$$f(x) = \sum_{m=1}^{M} \beta_m \overbrace{b(x; \gamma_m)}^{\text{Decision tree}},$$

Gradient Boosting Machines

$$(\beta_m, \gamma_m) = \arg\min_{\beta,\gamma} \sum_{i=1}^{N} L(y_i, f_{m-1}(x_i) + \beta b(x_i; \gamma)).$$

$$= \quad (y_i - f_{m-1}(x_i) - \beta b(x_i; \gamma))^2$$

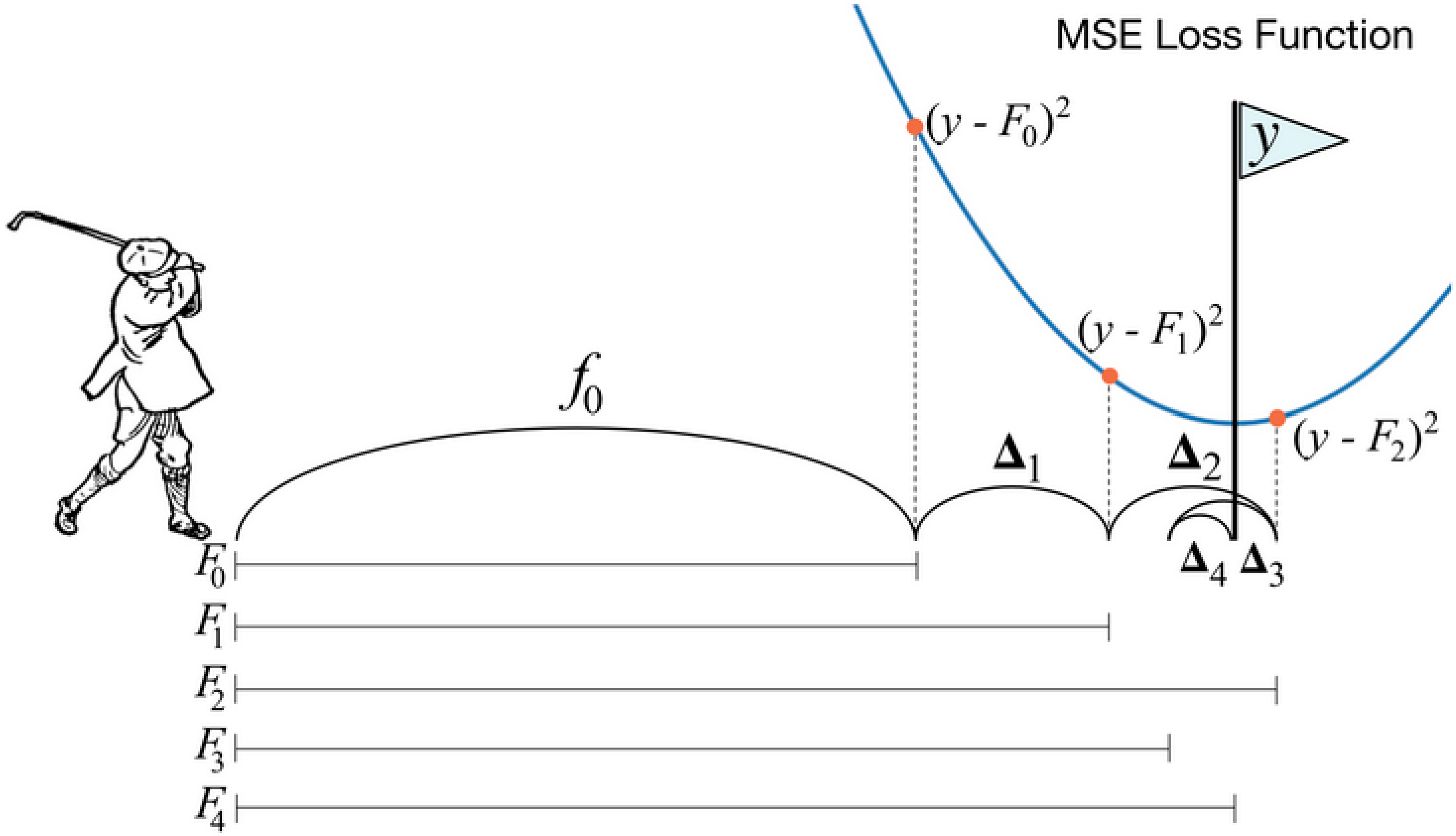
Gradient Boosting Machines

$$(\beta_m, \gamma_m) = \arg\min_{\beta,\gamma} \sum_{i=1}^{N} L(y_i, f_{m-1}(x_i) + \beta b(x_i; \gamma)).$$

$$\begin{aligned} &= (\underbrace{y_i - f_{m-1}}_{\textit{Residuals of previous classifier}}(x_i) - \beta b(x_i;\gamma))^2 \\ &= (r_{im} - \beta b(x_i;\gamma))^2, \end{aligned}$$

Each tree minimizes the difference between its predictions and the residuals of the previous tree.

Intuitively…

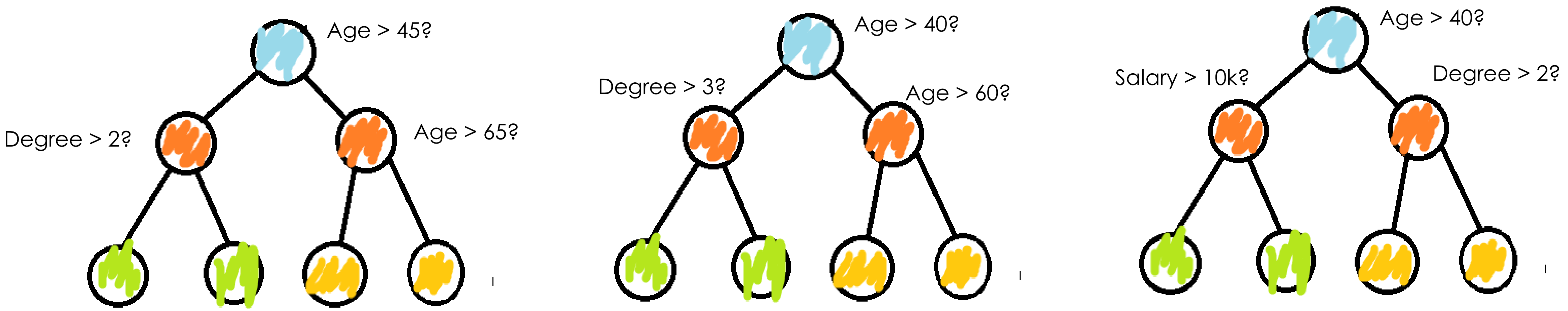


Each tree is nudging the approximation closer and closer to the real target value.

https://explained.ai/gradient-boosting/descent.html

Obtaining the prediction

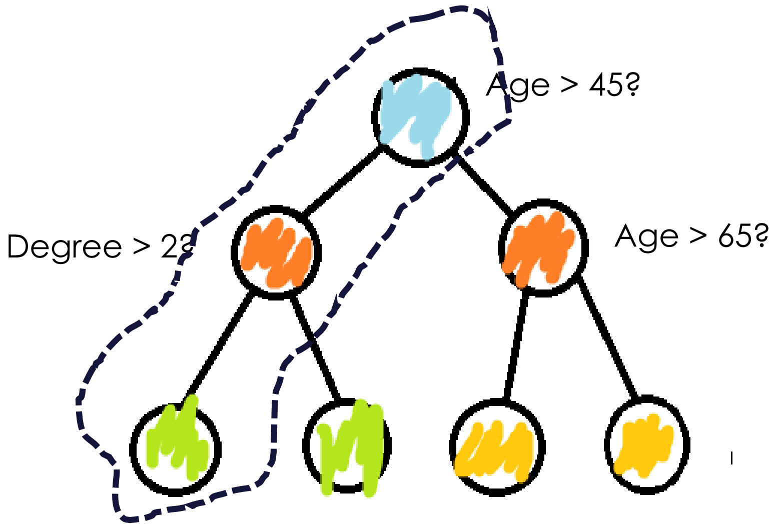
Boosting

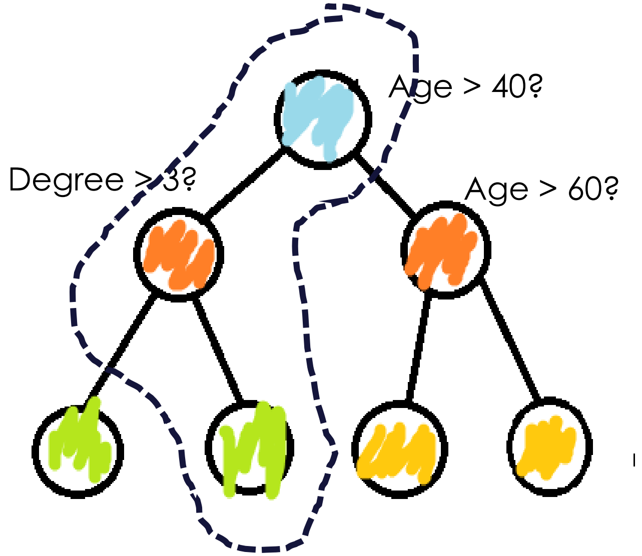


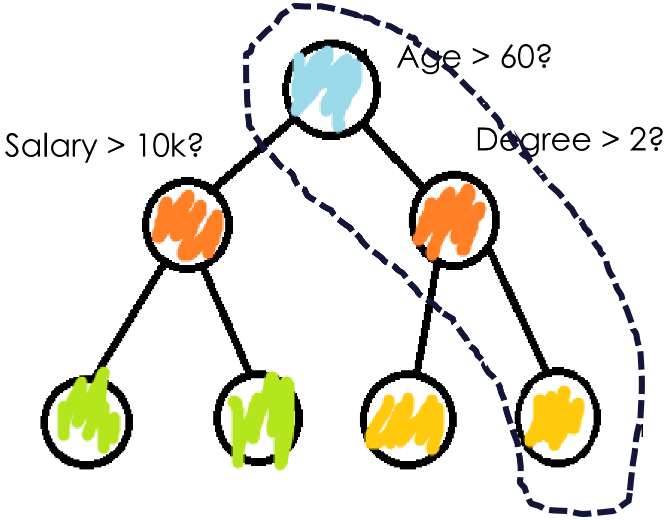
Age > 45?
Degree > 2?
Age > 65?
Age > 40?
Degree > 3?
Age > 60?
Age > 40?
Salary > 10k?
Degree > 2?

Boosting

Sample 1

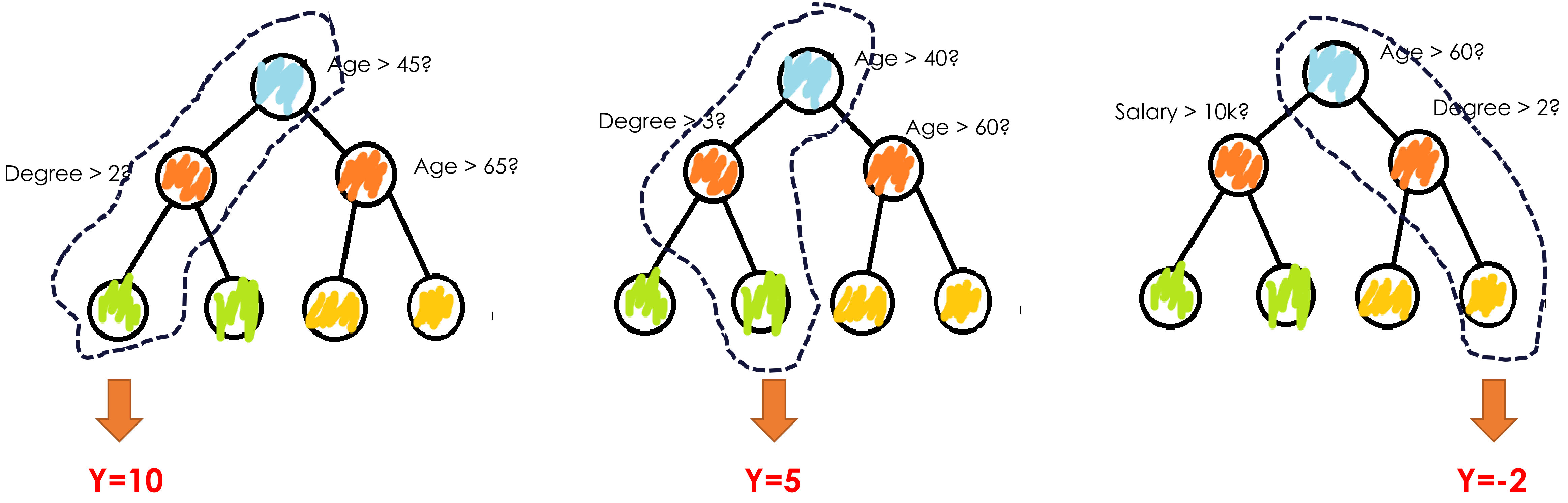






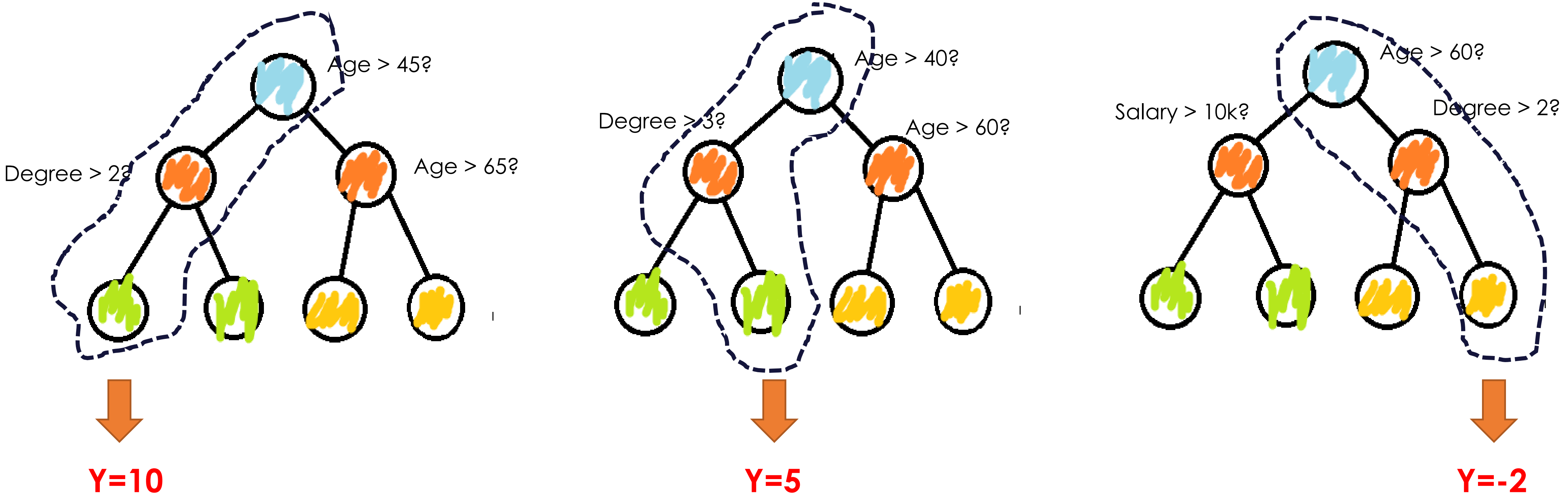
Boosting

Sample 1



Boosting

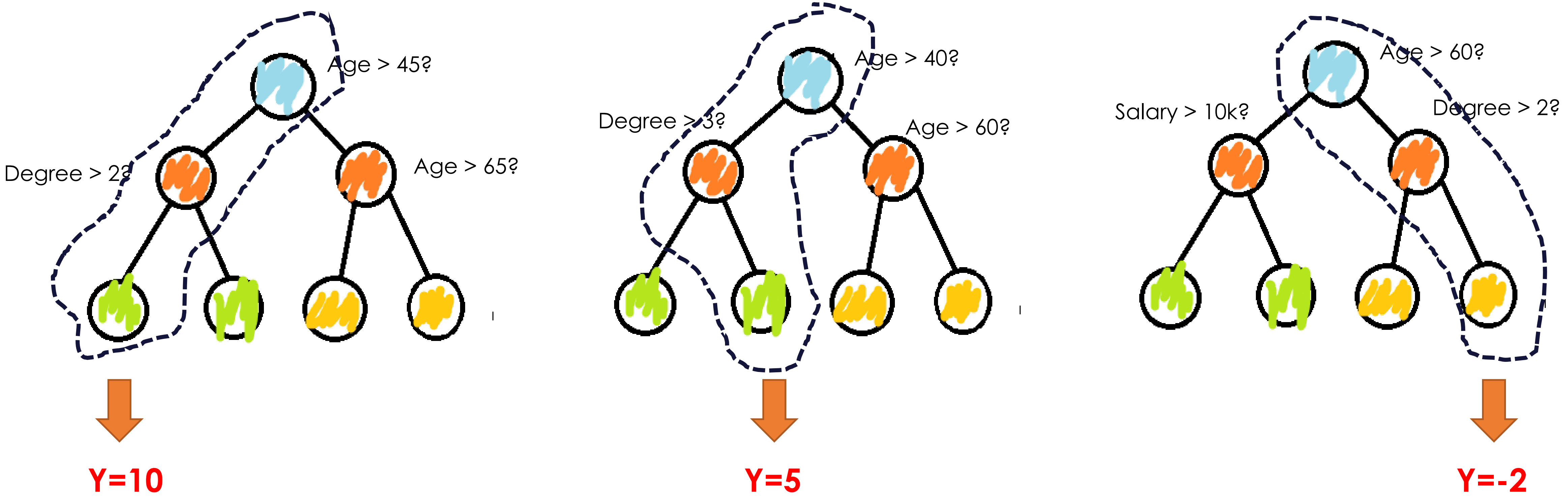
Sample 1



approx = 10 + 5 -2

Boosting

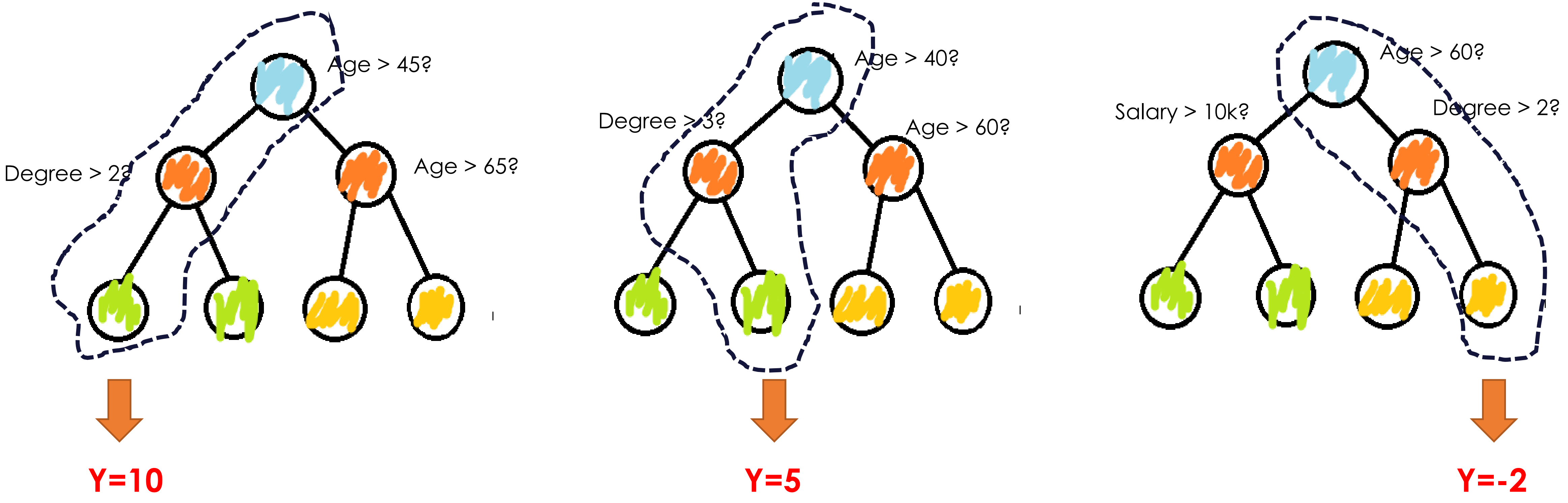
Sample 1



approx = w*10 + w*5 – w*2

Boosting

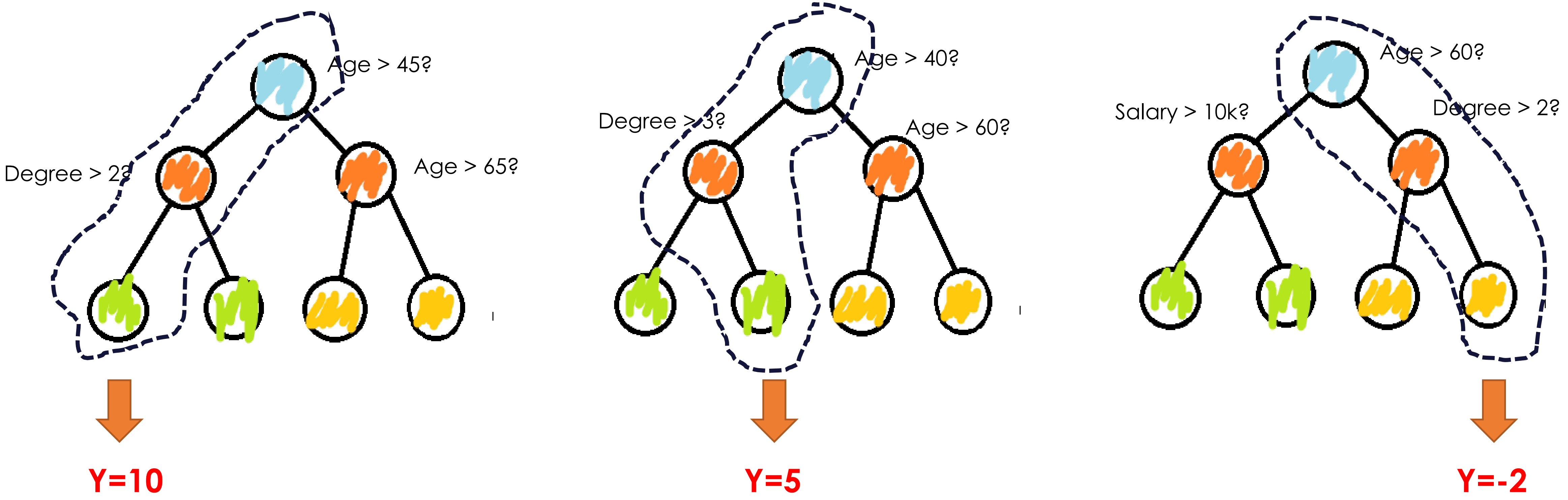
Sample 1



Prediction = **bias** + w*10 + w*5 – w*2

Boosting

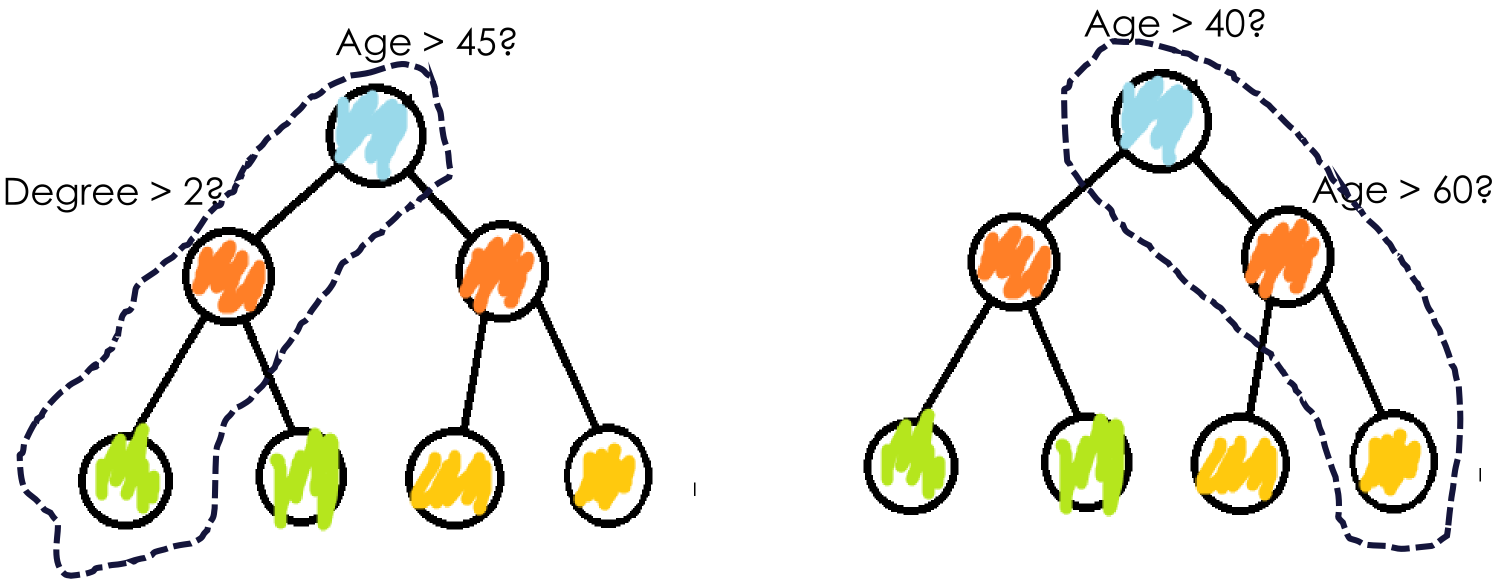
Sample 1



Prediction = **mean(y_train)** + w*10 + w*5 – w*2

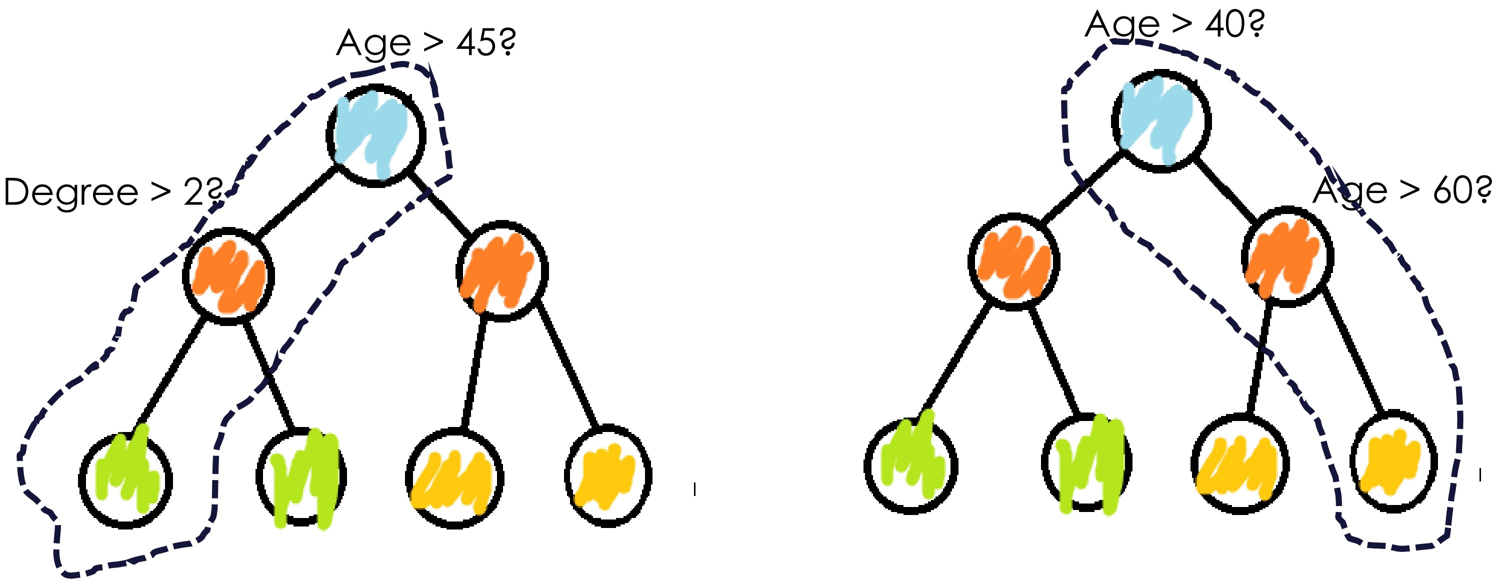
Obtaining feature contribution

Feature contribution



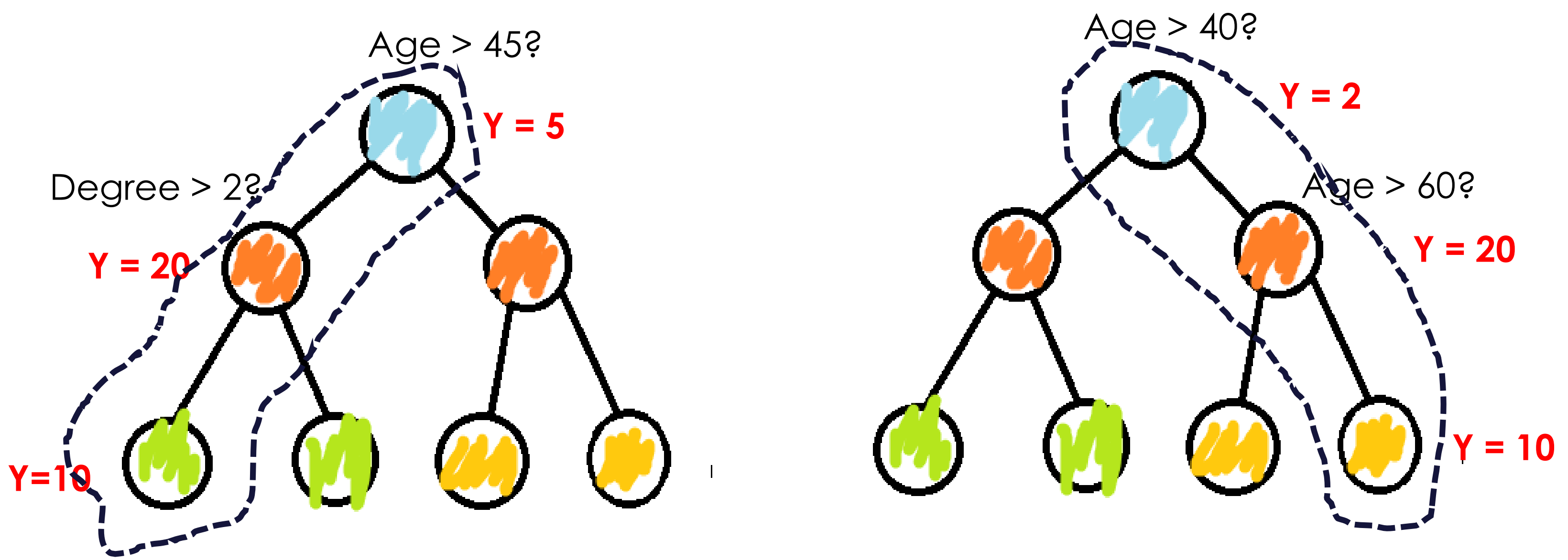
Each **split / feature** is also nudging the approximation closer and closer to the real target value.

Feature contribution



Feature contribution → its contribution at each split, aggregated over the ensemble.

Feature contribution



Age = 20 - 5 = 15

Degree = 10 - 20 = -10

+

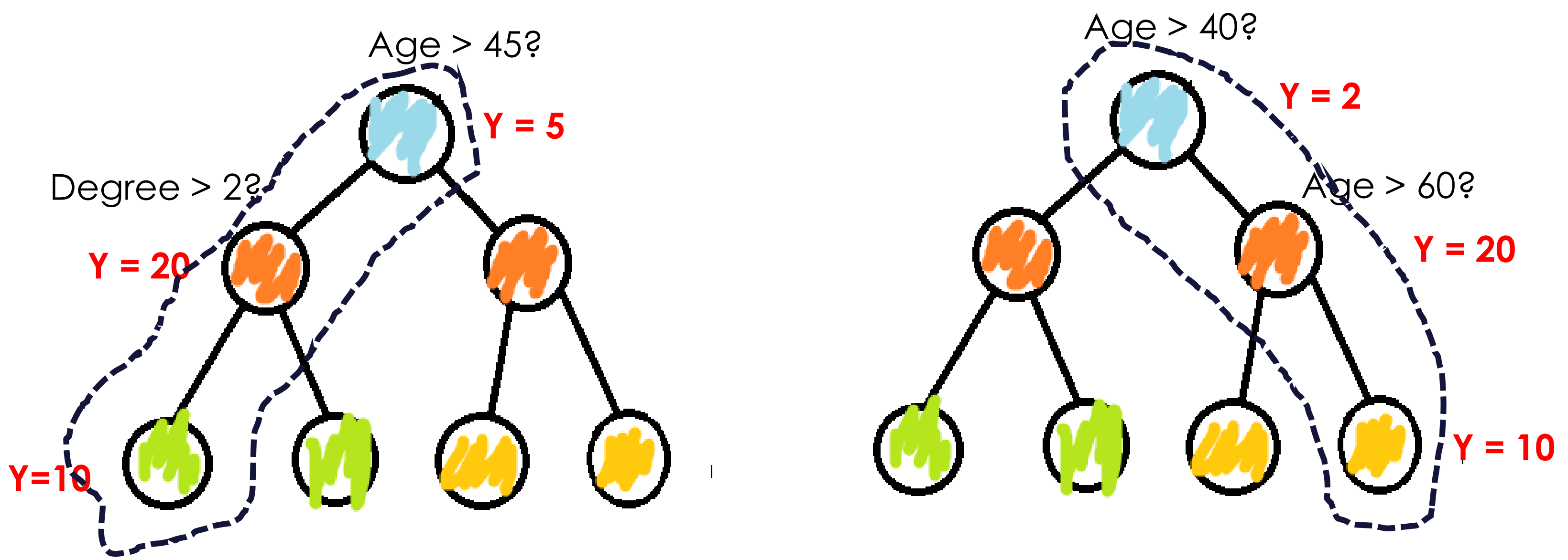
Age = (20 – 2) + (10-20) = 8

Degree = 0

Age = 15 + 8 = 23

Degree = -10 + 0 = -10

Feature contribution



Age = 20 - 5 = 15

Degree = 10 - 20 = -10

+

Age = (20 – 2) + (10-20) = 8

Degree = 0

Age = 15 + 8 = 23

Degree = -10 + 0 = -10

× 0.1

Learning rate = 0.1

THANK YOU

www.trainindata.com