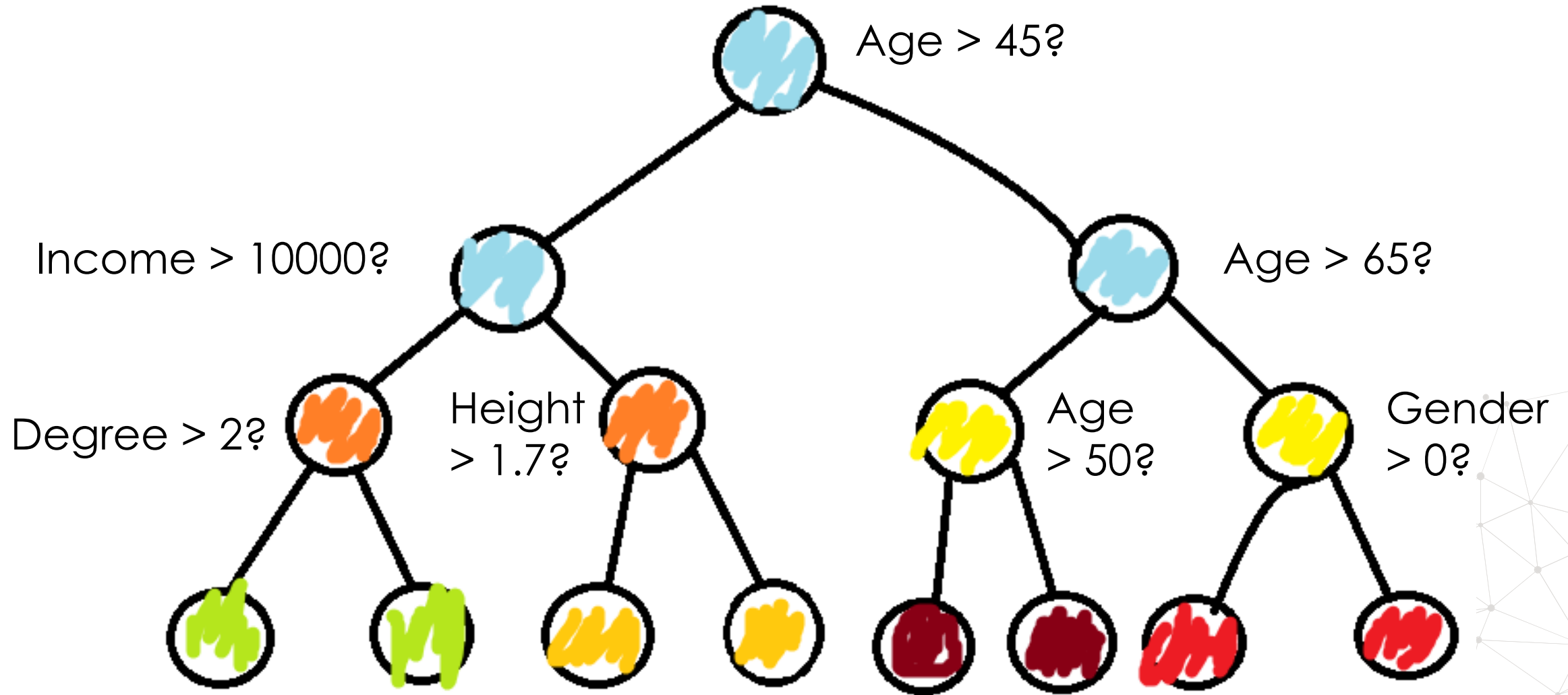




# Introducing Cost in Decision Trees

# Decision trees



# Decision Trees

$$G(Q, \theta) = \frac{n_{left}}{N_m} H(Q_{left}(\theta)) + \frac{n_{right}}{N_m} H(Q_{right}(\theta))$$

$$p_{mk} = 1/N_m \sum_{x_i \in R_m} I(y_i = k)$$

$p_{mk}$  = proportion of observations of class  $k$  at each node

$m$  = node

$k$  = class

$i$  = observation

<https://scikit-learn.org/stable/modules/tree.html#classification>

# Decision Trees

$$G(Q, \theta) = \frac{n_{left}}{N_m} H(Q_{left}(\theta)) + \frac{n_{right}}{N_m} H(Q_{right}(\theta))$$

$$p_{mk} = 1/N_m \sum_{x_i \in R_m} I(y_i = k)$$

Gini →  $H(X_m) = \sum_k p_{mk}(1 - p_{mk})$

Entropy →  $H(X_m) = - \sum_k p_{mk} \log(p_{mk})$

Misclassification →  $H(X_m) = 1 - \max(p_{mk})$

$p_{mk}$  = proportion of observations of class  $k$  at each node

$m$  = node

$k$  = class

$i$  = observation

<https://scikit-learn.org/stable/modules/tree.html#classification>

# Decision Trees

$$G(Q, \theta) = \frac{n_{left}}{N_m} H(Q_{left}(\theta)) + \frac{n_{right}}{N_m} H(Q_{right}(\theta))$$

$$p_{mk} = 1/N_m \sum_{x_i \in R_m} I(y_i = k)$$

Gini →  $H(X_m) = \sum_k p_{mk}(1 - p_{mk})$  Wk

Entropy →  $H(X_m) = - \sum_k p_{mk} \log(p_{mk})$  Wk

Misclassification →  $H(X_m) = 1 - \max(p_{mk})$  Wk

$p_{mk}$  = proportion of observations of class k at each node

m = node

k = class

i = observation

<https://scikit-learn.org/stable/modules/tree.html#classification>



# Tree ensembles

Ensemble of trees like Random optimise the same functions, so the logic is the same



# THANK YOU

[www.trainindata.com](http://www.trainindata.com)