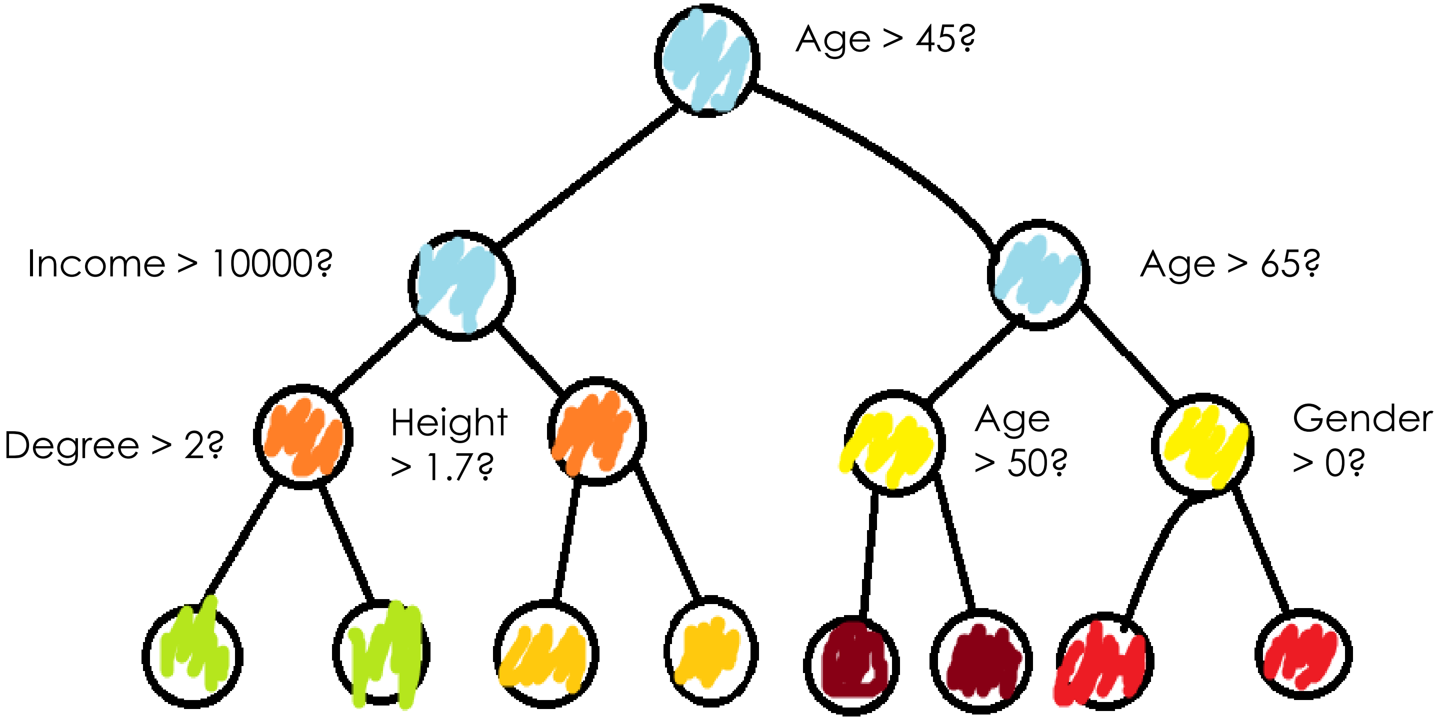
Decision trees induction

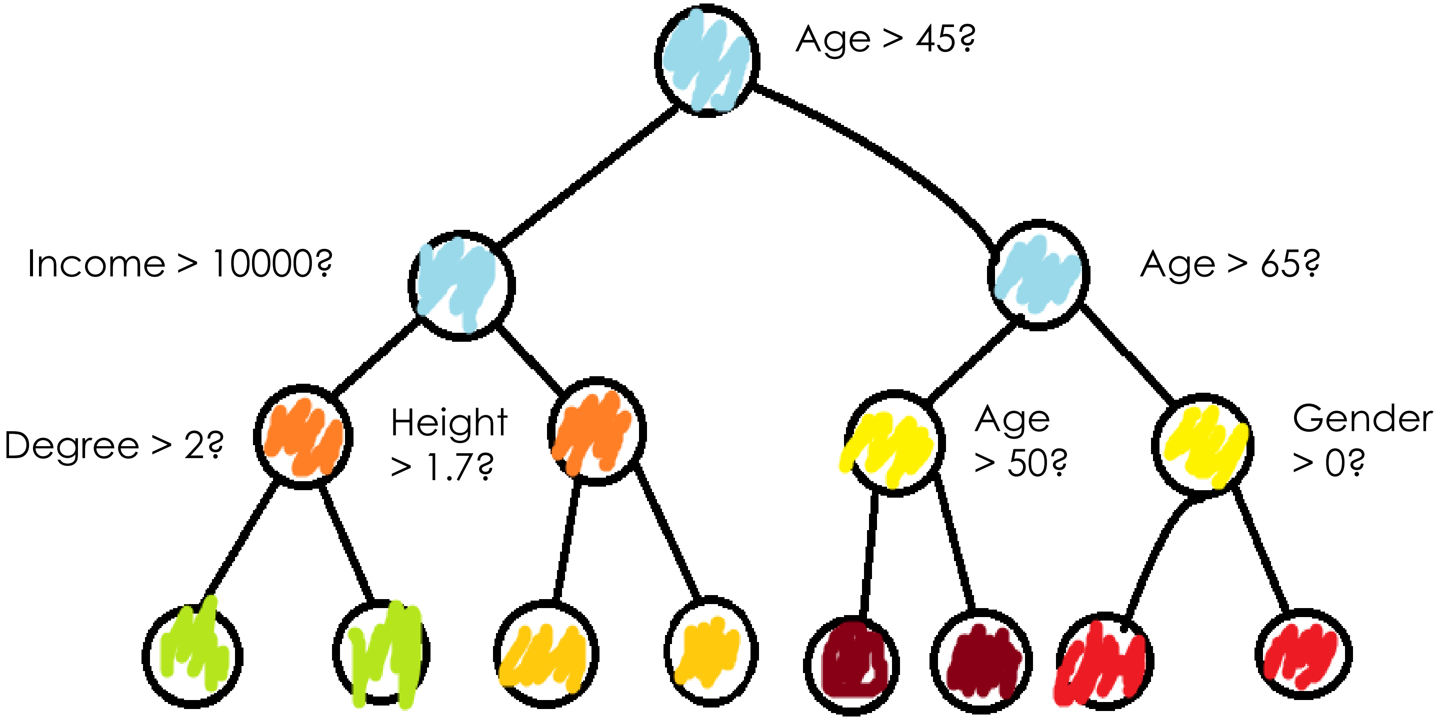
Decision trees: induction



At each node:

Choose variable and split point that achieve best fit.

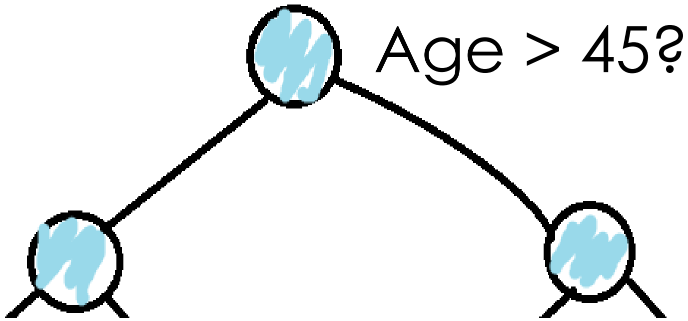
Decision trees: induction



At each node:

Best split is quantified with a function.

Decision trees: induction

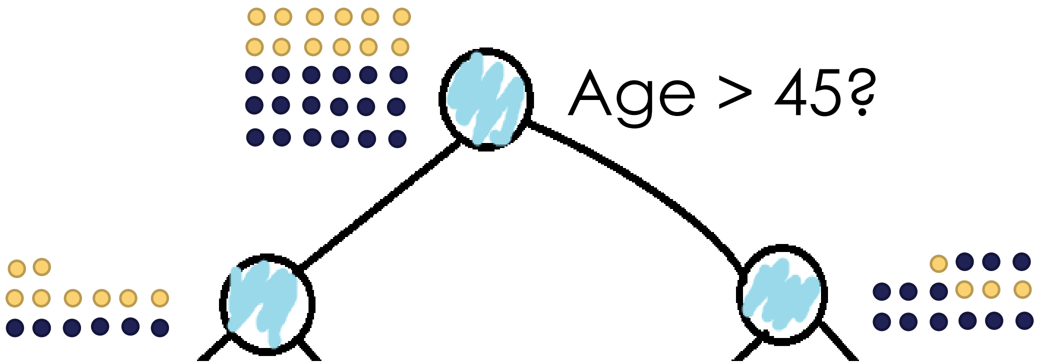


$$G = \frac{n_{left}}{N_m} H_{left} + \frac{n_{right}}{N_m} H_{right}$$

G is the impurity (loss) function.

Two terms: one from the left child, and one from the right child.

Decision trees: induction



$$G = \frac{n_{left}}{N_m} H_{left} + \frac{n_{right}}{N_m} H_{right}$$

G is the impurity (loss) function.

Weighted by the fraction of samples at each child.

Decision trees: classification

$$G = \frac{n_{left}}{N_m} H_{left} + \frac{n_{right}}{N_m} H_{right}$$

pmk = proportion of observations of class k at each node

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

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

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

Decision trees: regression

$$G = \frac{n_{left}}{N_m} H_{left} + \frac{n_{right}}{N_m} H_{right}$$

Sum of squares ➔

$$H(Q_m) = \frac{1}{n_m} \sum_{y \in Q_m} (y - \bar{y}_m)^2$$

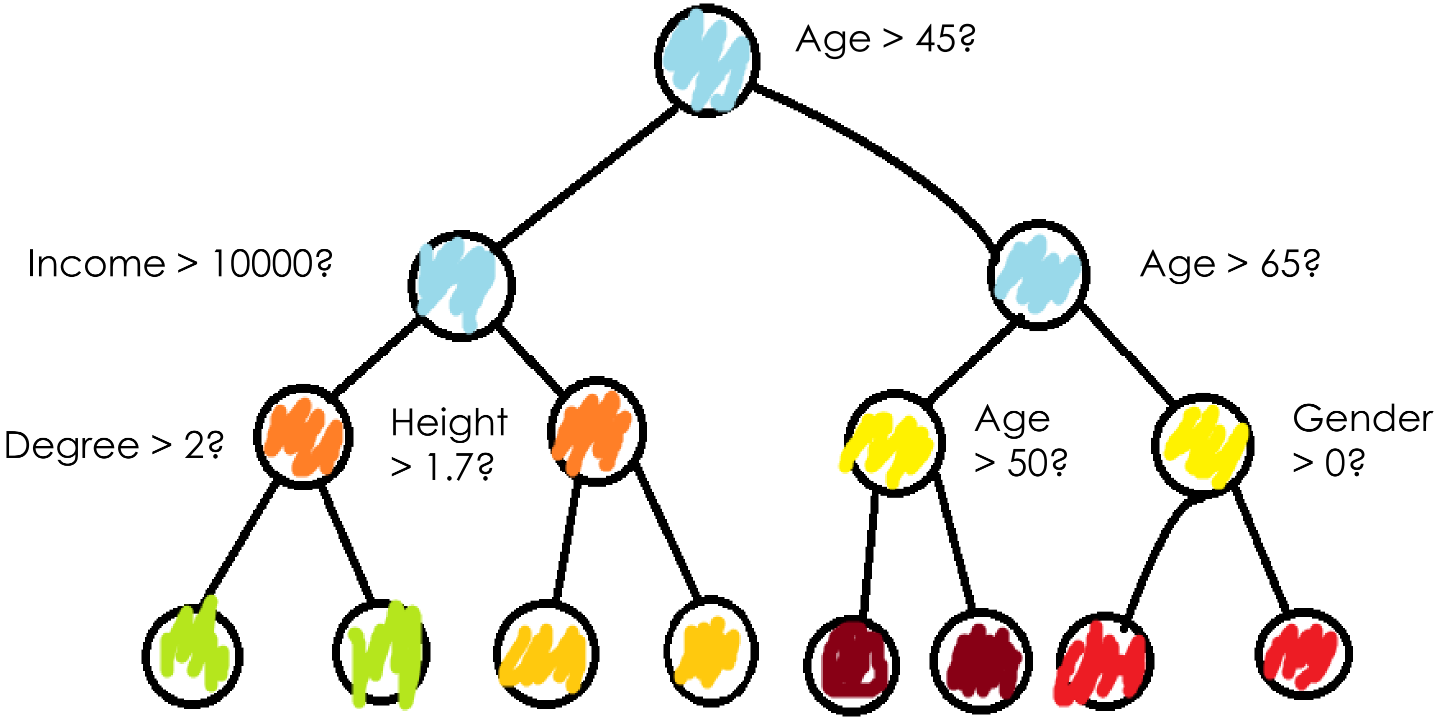
Poisson deviance ➔

$$H(Q_m) = \frac{1}{n_m} \sum_{y \in Q_m} \left(y \log \frac{y}{\bar{y}_m} - y + \bar{y}_m\right)$$

Mean absolute error ➔

$$H(Q_m) = \frac{1}{n_m} \sum_{y \in Q_m} |y - median(y)_m|$$

Decision trees: induction



Process:

Continues until stopping criteria.

Loss function ➔ feature importance

$$G = \frac{n_{left}}{N_m} H_{left} + \frac{n_{right}}{N_m} H_{right}$$

The loss function is used to infer feature importance.

THANK YOU

www.trainindata.com