Inductive Bias

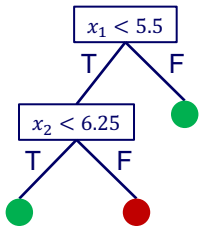
Mark Craven and David Page
Computer Sciences 760
Spring 2019

Inductive bias

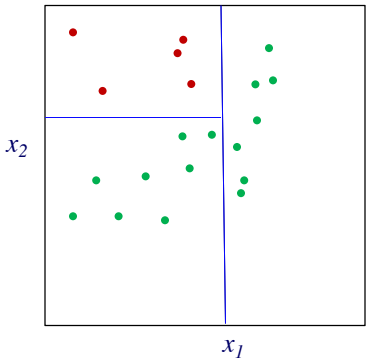
- *inductive bias* is the set of assumptions a learner uses to be able to predict y for a previously unseen instance $\boldsymbol{x}$
- two components
  - *hypothesis space bias*: determines the models that can be represented
  - *preference bias*: specifies a preference ordering within the space of models
- in order to *generalize* (i.e. make predictions for previously unseen instances) a learning algorithm must have an inductive bias

Decision tree hypothesis space bias

- decision trees partition the instance space with axis-parallel splits

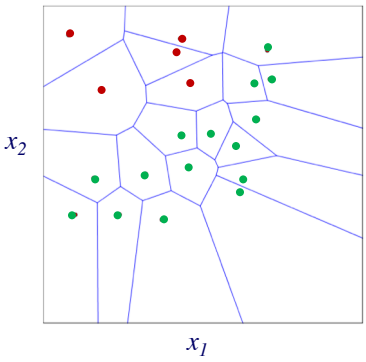


axis-parallel splits

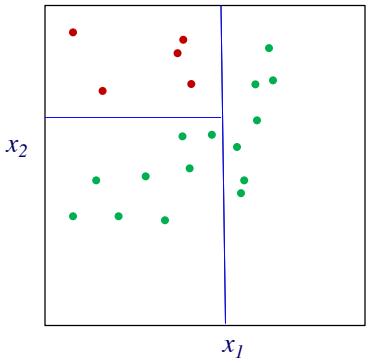


Hypothesis space bias: *k*-NN versus decision tree

Voronoi tessellation



axis-parallel splits



Preference bias: ID3 vs. C4.5 vs. CART

- All three methods have the same hypothesis space bias for classification problems, but differ in their preference bias

	ID3	C4.5	CART
split selection criterion	information gain	gain ratio	Gini impurity
split types	*n*-way for discrete features with *n* values	*n*-way for discrete features with *n* values	binary
overfitting avoidance	early stopping	Reduced-error pruning	cost-complexity pruning