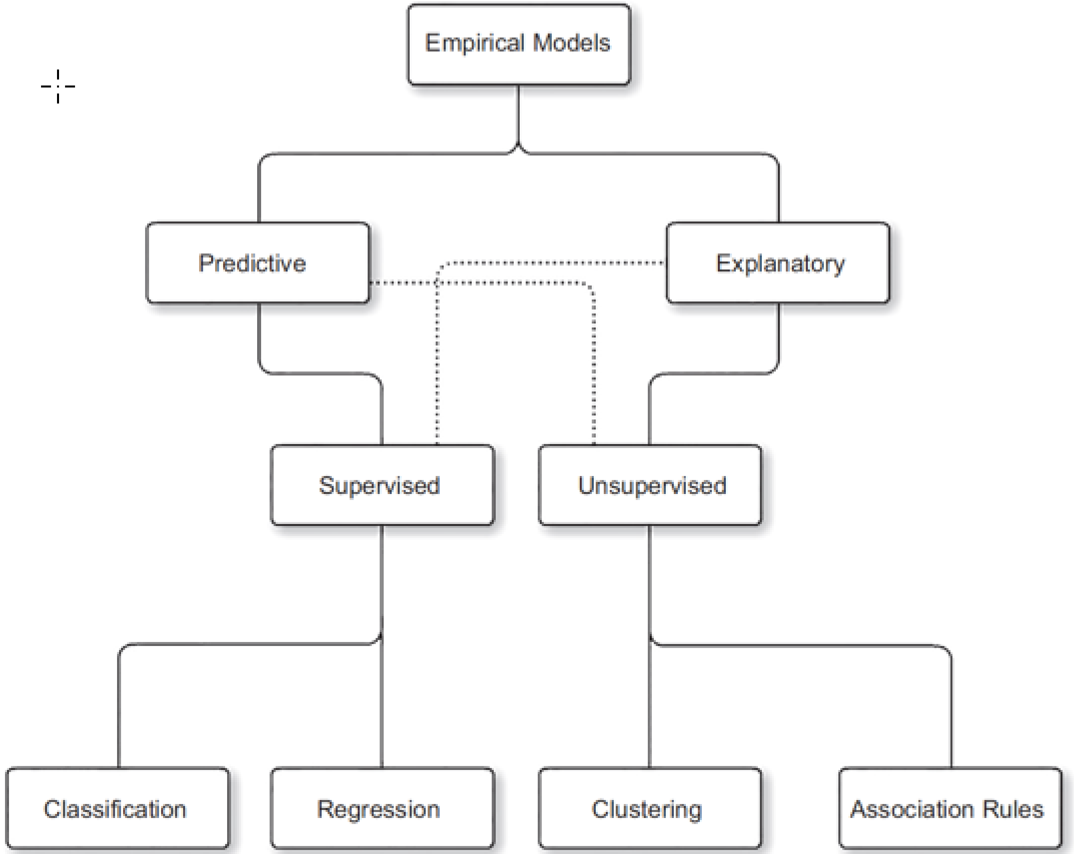
4. Classification

Models



Empirical Models
Predictive
Explanatory
Supervised
Unsupervised
Classification
Regression
Clustering
Association Rules

Classification

Target variable is categorical. Predictors could be of any data type.

Algorithms

Decision Trees

Rule induction

kNN

Naive Bayesian

Neural Networks

Support Vector Machines

Ensemble Meta Models

Decision Trees

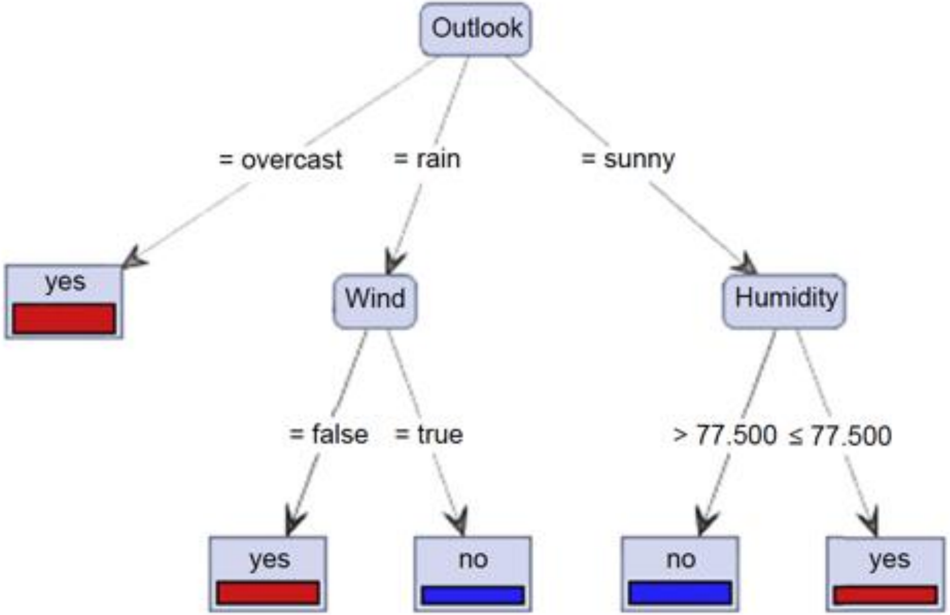
Decision Trees

Predictors / Attributes **Target / Class**

Outlook	Temperature	Humidity	Windy	Play
sunny	85	85	FALSE	no
sunny	80	90	TRUE	no
overcast	83	78	FALSE	yes
rain	70	96	FALSE	yes
rain	68	80	FALSE	yes
rain	65	70	TRUE	no
overcast	64	65	TRUE	yes
sunny	72	95	FALSE	no
sunny	69	70	FALSE	yes
rain	75	80	FALSE	yes
sunny	75	70	TRUE	yes
overcast	72	90	TRUE	yes
overcast	81	75	FALSE	yes
rain	71	80	TRUE	no

Decision Tree

Outlook	Temperature	Humidity	Windy	Play
sunny	85	85	FALSE	no
sunny	80	90	TRUE	no
overcast	83	78	FALSE	yes
rain	70	96	FALSE	yes
rain	68	80	FALSE	yes
rain	65	70	TRUE	no
overcast	64	65	TRUE	yes
sunny	72	95	FALSE	no
sunny	69	70	FALSE	yes
rain	75	80	FALSE	yes
sunny	75	70	TRUE	yes
overcast	72	90	TRUE	yes
overcast	81	75	FALSE	yes
rain	71	80	TRUE	no



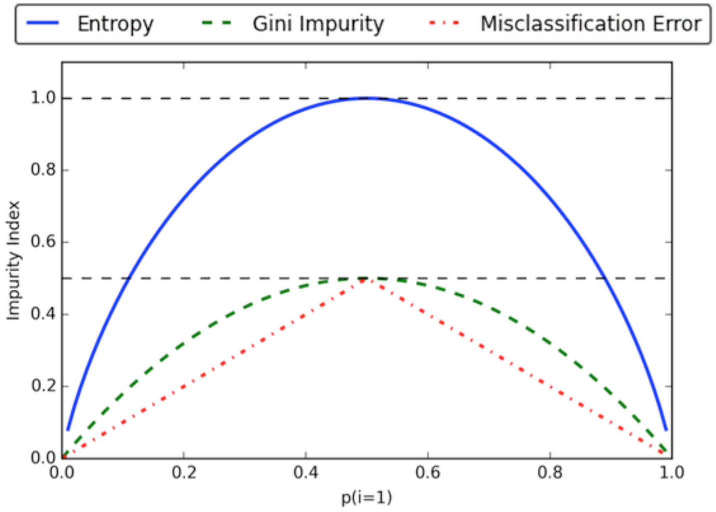
Measure of impurity

Every split ties to make child node more pure.

Gini impurity

Information Gain (Entropy)

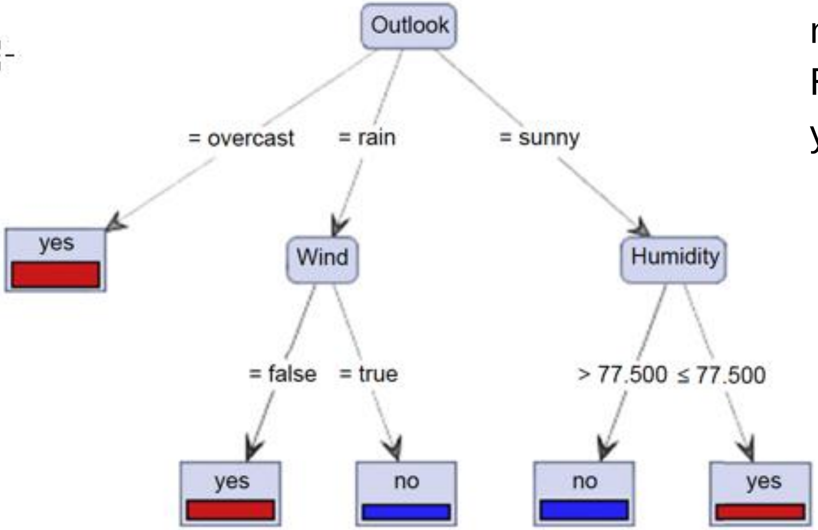
Misclassification Error



https://www.quora.com/What-are-the-advantages-of-different-Decision-Trees-Algorithms

Rule Induction

Tree to Rules



Rule 1: if (Outlook = overcast) then yes

Rule 2: if (Outlook = rain) and (Wind = false) then yes

Rule 3: if (Outlook = rain) and (Wind = true) then no

Rule 4: if (Outlook = sunny and (Humidity > 77.5) then no

Rule 5: if (Outlook = sunny and (Humidity ≤ 77.5) then yes

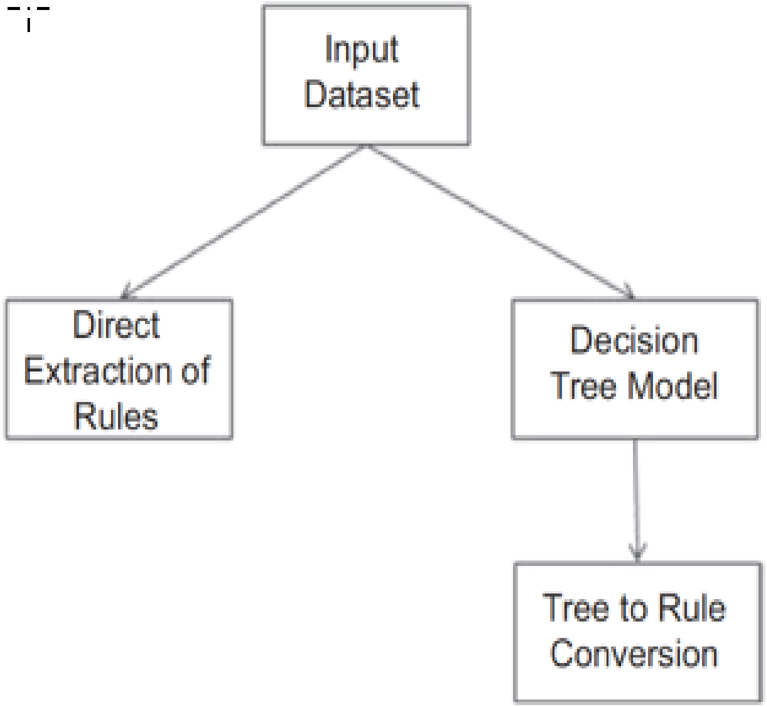
Rules

$R = \{ r_1 \cap r_2 \cap r_3 \cap .. r_k \}$

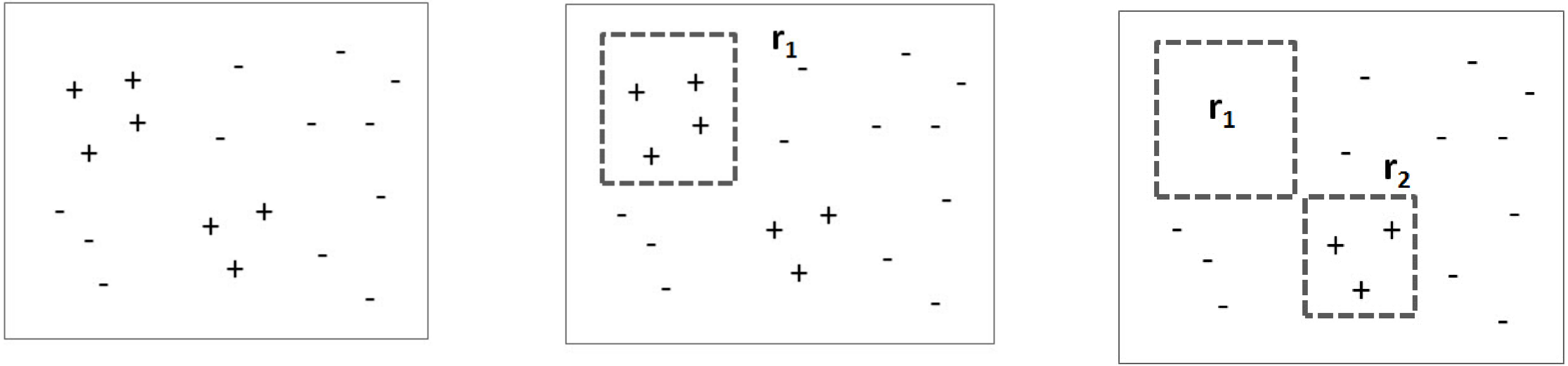
Where k is the number of disjuncts in a rule set. Individual disjuncts can be represented as

r_i = (antecedent or condition) then (consequent)

Approaches


Input Dataset
Direct Extraction of Rules
Decision Tree Model
Tree to Rule Conversion

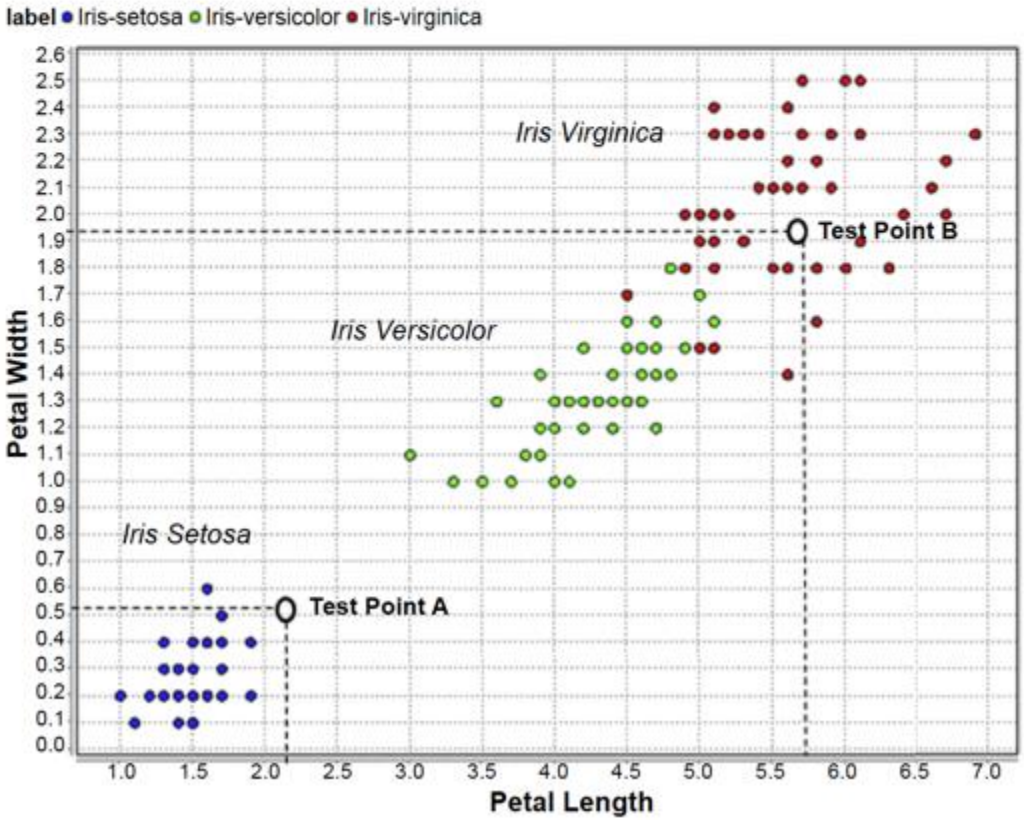
Sequential covering



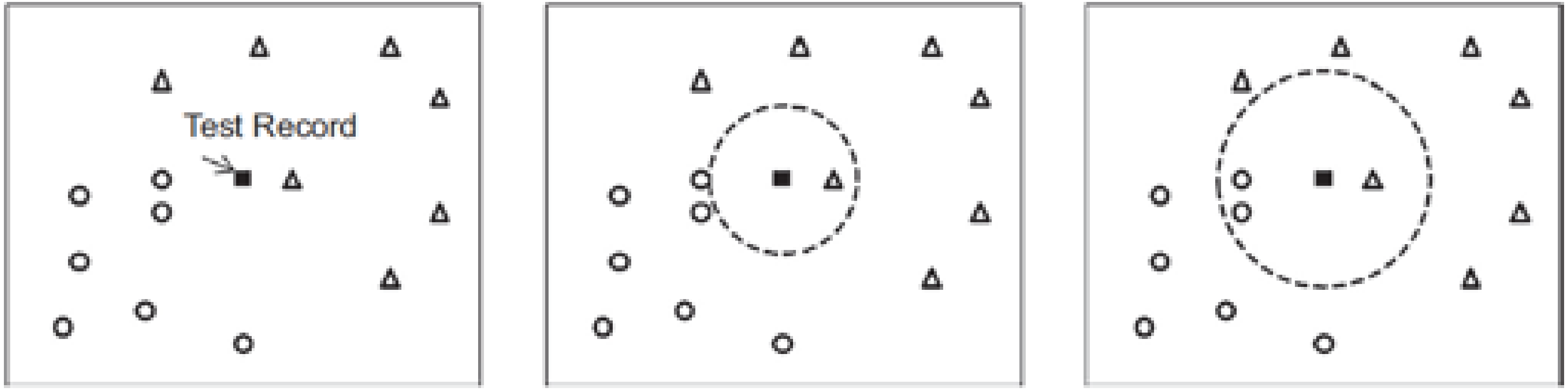
$$\text{Rule accuracy } A(r_i) = \frac{\text{Correct records covered by rule}}{\text{All records covered by the rule}}$$

K Nearest Neighbors

Guess the species for A and B



KNN



K=1 Predicted Class is triangle

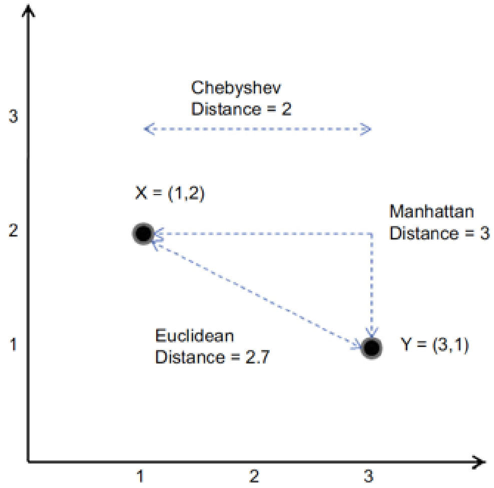
K=3 Predicted Class is circle

Measure of Proximity

Distance

$$\text{Distance } d = \sqrt{(x1 - y1)^2 + (x2 - y2)^2}$$

$$\text{Distance } d = \sqrt{(x1 - y1)^2 + (x2 - y2)^2 + \cdots + (xn - yn)^2}$$



Measure of Proximity

Correlation similarity

$$\text{Correlation }(X, Y) = \frac{s_{xy}}{s_x * s_y}$$

Simple matching coefficient

$$\text{Simple matching coefficient (SMC)} = \frac{\textit{matching occurences}}{\textit{total occurences}}$$

Jaccard Similarity

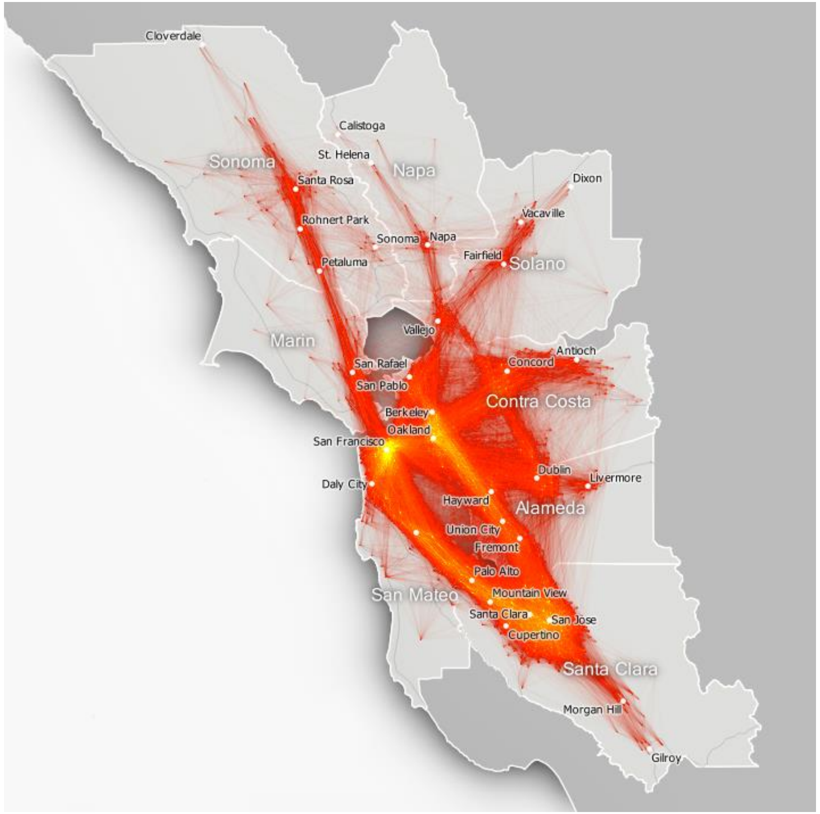
$$\text{Jaccard coefficient} = \frac{\textit{common occurences}}{\textit{total occurences}}$$

Cosine similarity

$$\text{Cosine similarity }(|X, Y|) = \frac{x \cdot y}{\| x \| \| y \|}$$

NAÏVE BAYESIAN

Predict your commute time



http://www.wired.com/2015/08/pretty-maps-bay-area-hellish-commutes/#slide-2

Bayes' theorem

Probability of the outcome

Class conditional probability

$$P(Y|\mathbf{X}) = \frac{P(Y) * P(\mathbf{X}|Y)}{P(\mathbf{X})}$$

Posterior probability

Probability of the conditions

Data set

Table 4.4 Golf Data Set with Modified Temperature and Humidity Attributes

No.	Temperature X_1	Humidity X_2	Outlook X_3	Wind X_4	Play (Class Label) Y
1	high	med	sunny	false	no
2	high	high	sunny	true	no
3	low	low	rain	true	no
4	med	high	sunny	false	no
5	low	med	rain	true	no
6	high	med	overcast	false	yes
7	low	high	rain	false	yes
8	low	med	rain	false	yes
9	low	low	overcast	true	yes
10	low	low	sunny	false	yes
11	med	med	rain	false	yes
12	med	low	sunny	true	yes
13	med	high	overcast	true	yes
14	high	low	overcast	false	yes

Class conditional probability

Class Conditional Probability of Temperature		
Temperature (X_1)	**$P(X_1 \mid Y = no)$**	**$P(X_1 \mid Y = yes)$**
high	2/5	2/9
med	1/5	3/9
low	2/5	4/9

Conditional Probability of Humidity, Outlook, and Wind		
Humidity (X_2)	**$P(X_1 \mid Y = no)$**	**$P(X_1 \mid Y = yes)$**
high	2/5	2/9
low	1/5	4/9
med	2/5	3/9
Outlook (X_3)	**$P(X_1 \mid Y = no)$**	**$P(X_1 \mid Y = yes)$**
overcast	0/5	4/9
Rain	2/5	3/9
sunny	3/5	2/9
Wind (X_4)	**$P(X_1 \mid Y = no)$**	**$P(X_1 \mid Y = yes)$**
false	2/5	6/9
true	3/5	3/9

Test record

Table 4.7 Test Record

No.	Temperature X_1	Humidity X_2	Outlook X_3	Wind X_4	Play (Class Label) Y
Unlabeled Test	high	low	sunny	false	?

Calculation of posterior probability P(Y/X)

$$P(Y = yes|\mathbf{X}) = \frac{P(Y) * \prod_{i=1}^{n} P(Xi \mid Y)}{P(X)}$$

$= P(Y = yes) * \{P(Temp = high|Y = yes) * P(Humidity = low|Y = yes) * P(Outlook = sunny|\ Y = yes) * P(Wind = false|Y = yes)\}/P(X)$

$= 9/14 * \{2/9 * 4/9 * 2/9 * 6/9\}/P(X)$

$= 0.0094/P(X)$

$P(Y = no|\mathbf{X}) = 5/14 * \{2/5 * 4/5 * 3/5 * 2/5\}$

$= 0.0274/P(X)$

We normalize both the estimates by dividing both by (0.0094 + 0.027) to get

$$\text{Likelihood of } (Play = yes) = \frac{0.0094}{0.0274 + 0.0094} = 26\%$$

$$\text{Likelihood of } (Play = no) = \frac{0.0094}{0.0274 + 0.0094} = 74\%$$

Issues

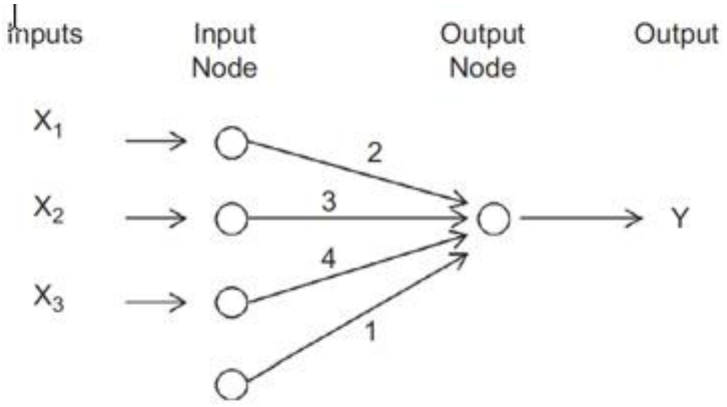
Incomplete training set -> Use laplace correction

Continuous numeric attributes -> Use Probability density function

Attributes independence -> remove correlated attributes

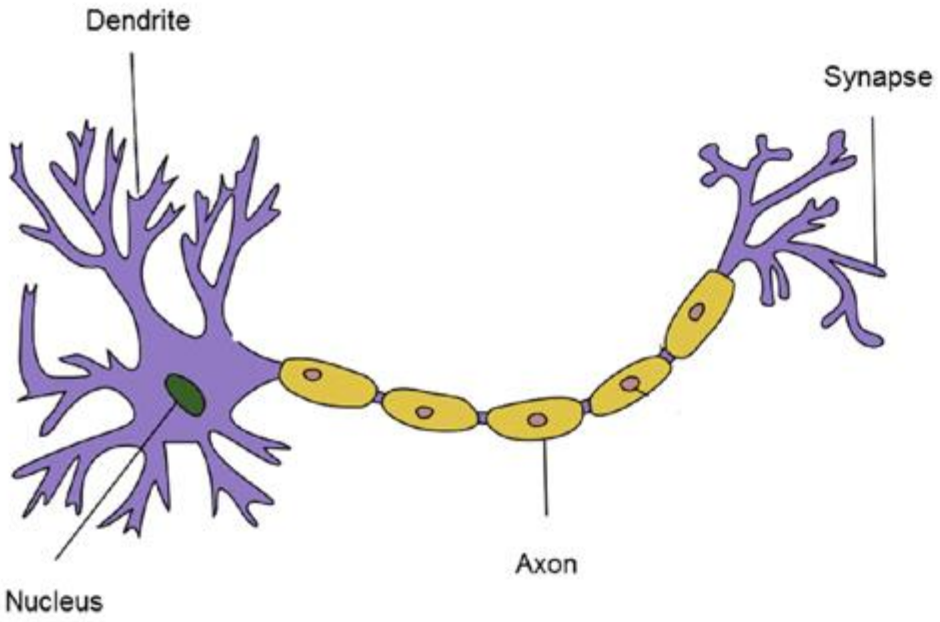
NEURAL NETWORKS

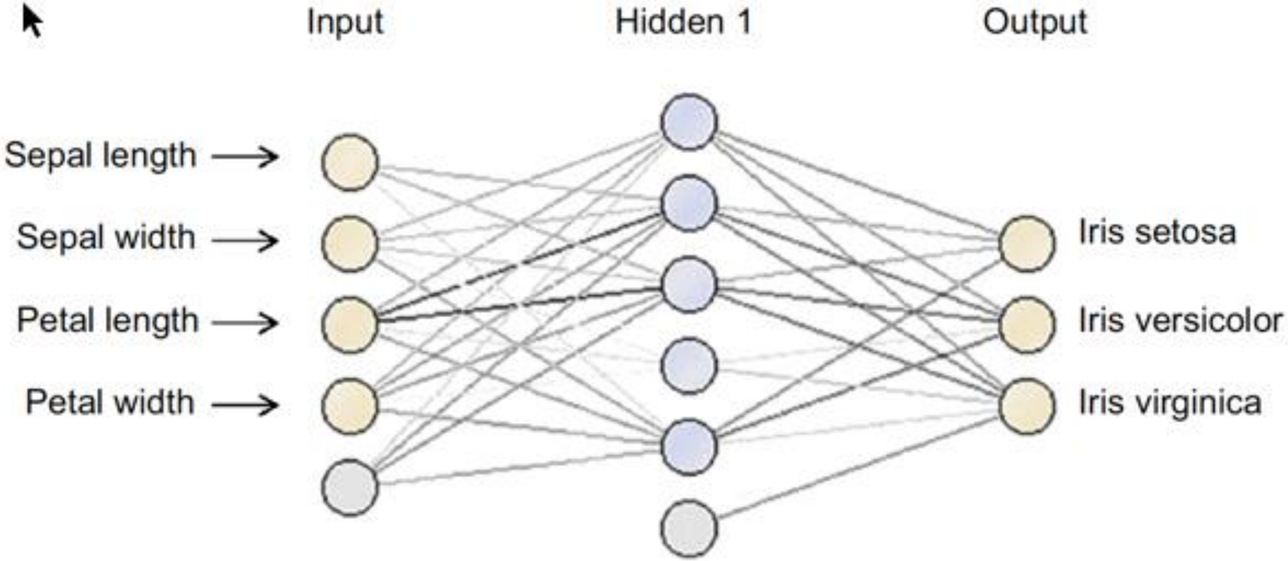
Model

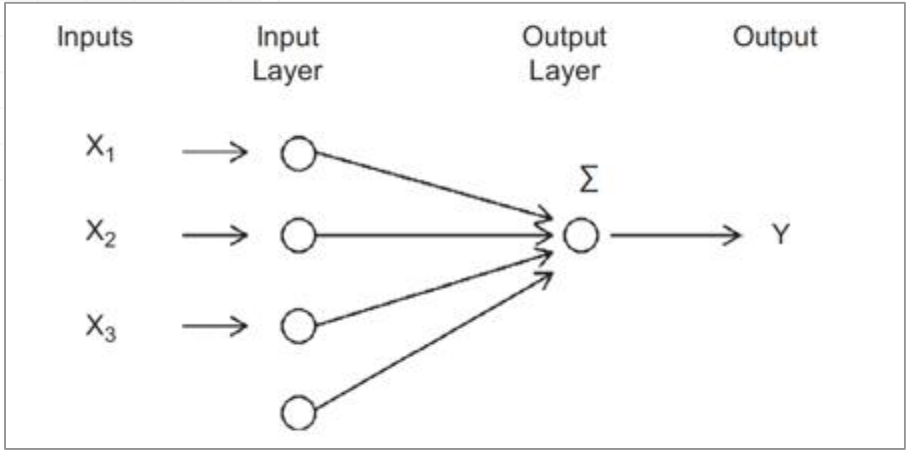


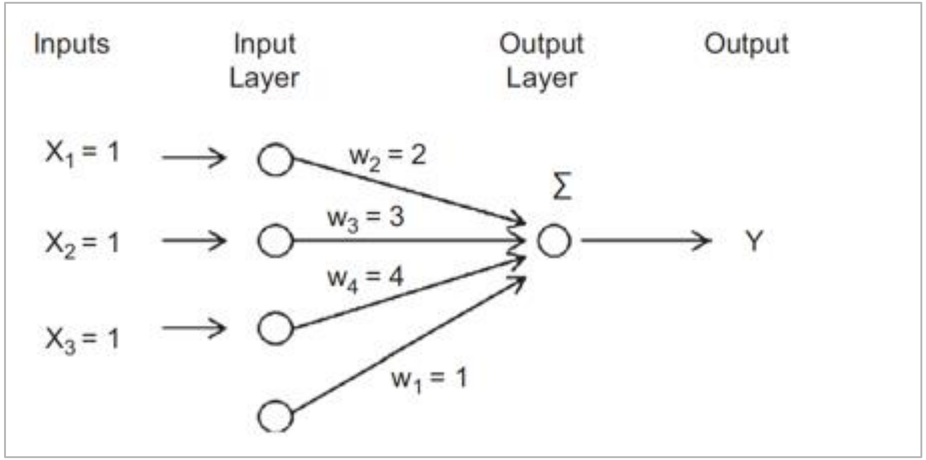
Y = 1 + 2X1 + 3X2 + 4X3

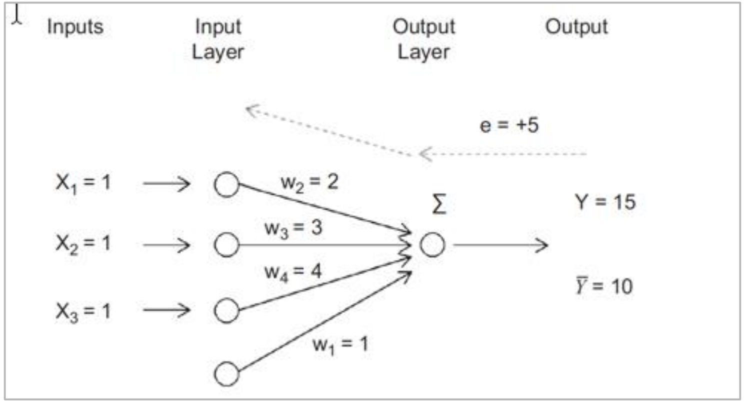
Neurons




Input
Hidden 1
Output
Sepal length
Sepal width
Petal length
Petal width
Iris setosa
Iris versicolor
Iris virginica

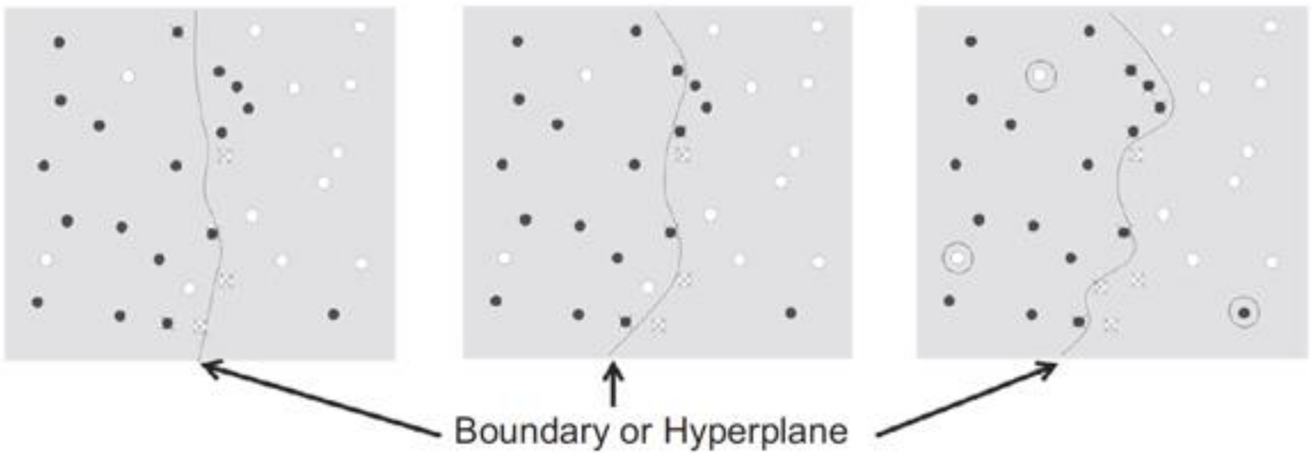

Inputs
Input Layer
Output Layer
Output
X1
X2
X3
Σ
Y


Inputs
Input Layer
Output Layer
Output
X1 = 1
X2 = 1
X3 = 1
w2 = 2
w3 = 3
w4 = 4
w1 = 1
Σ
Y

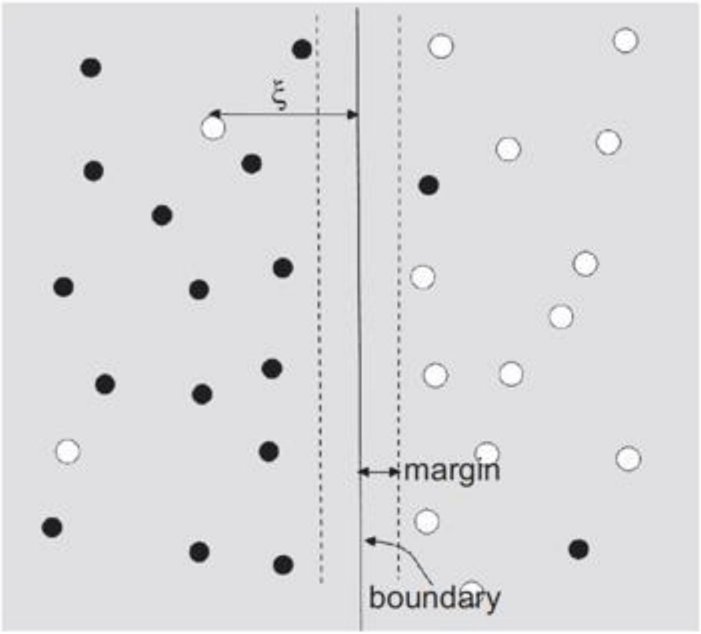

Inputs
Input Layer
Output Layer
Output
e = +5
X1 = 1
X2 = 1
X3 = 1
w2 = 2
w3 = 3
w4 = 4
w1 = 1
Σ
Y = 15
Ȳ = 10

SUPPORT VECTOR MACHINES

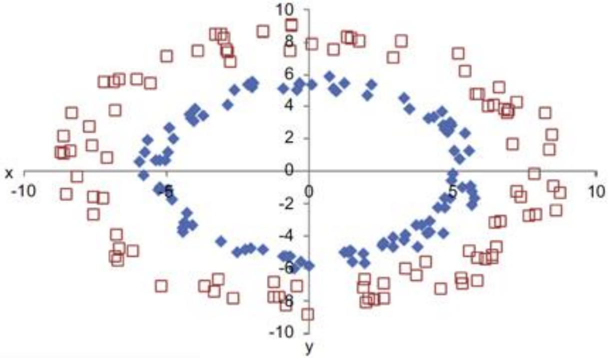
Boundary

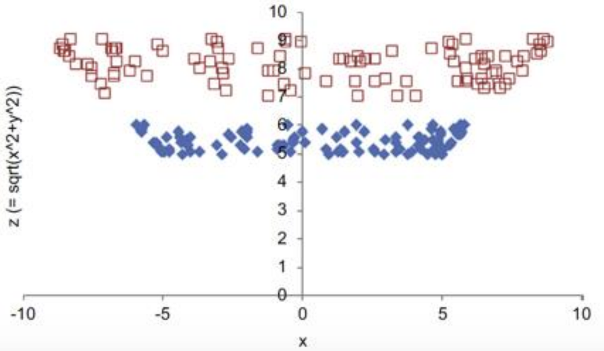


Margin

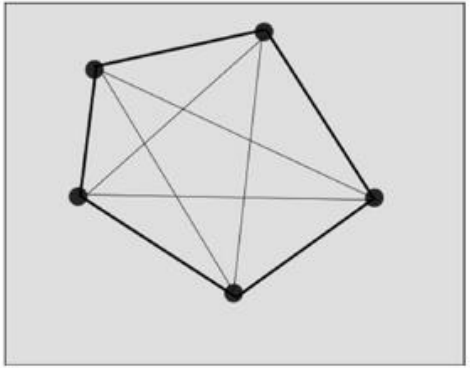


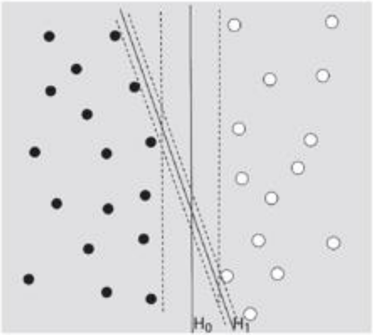
Transforming linearly non-separable data





Optimal hyperplane





Ensemble Learners

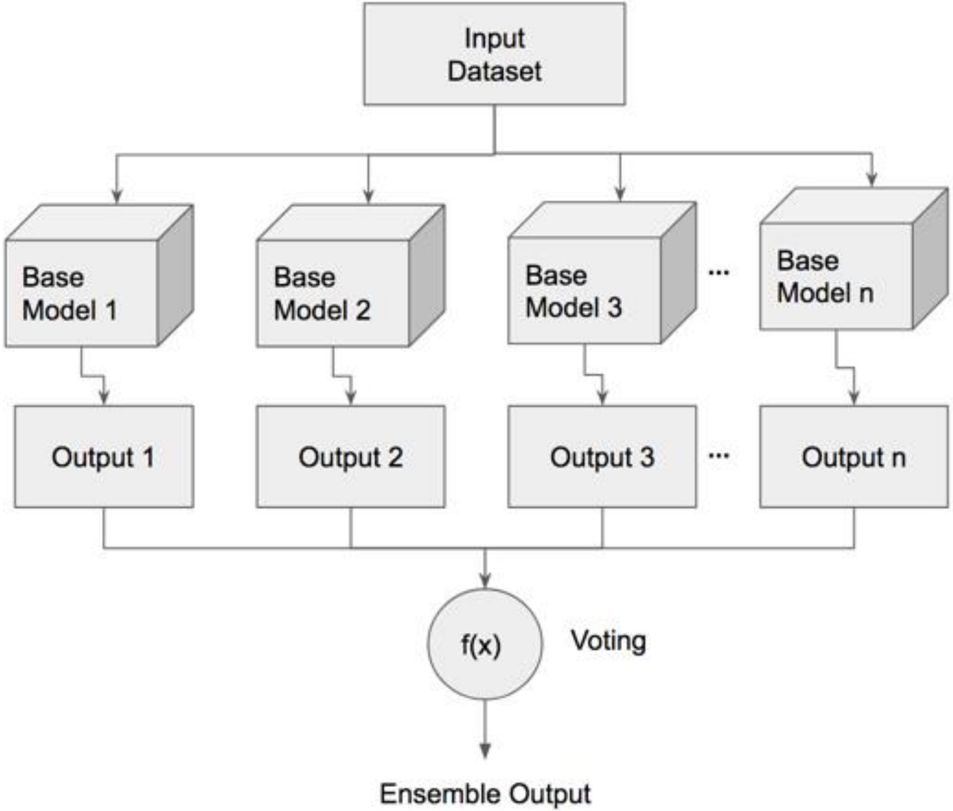
Ensemble model

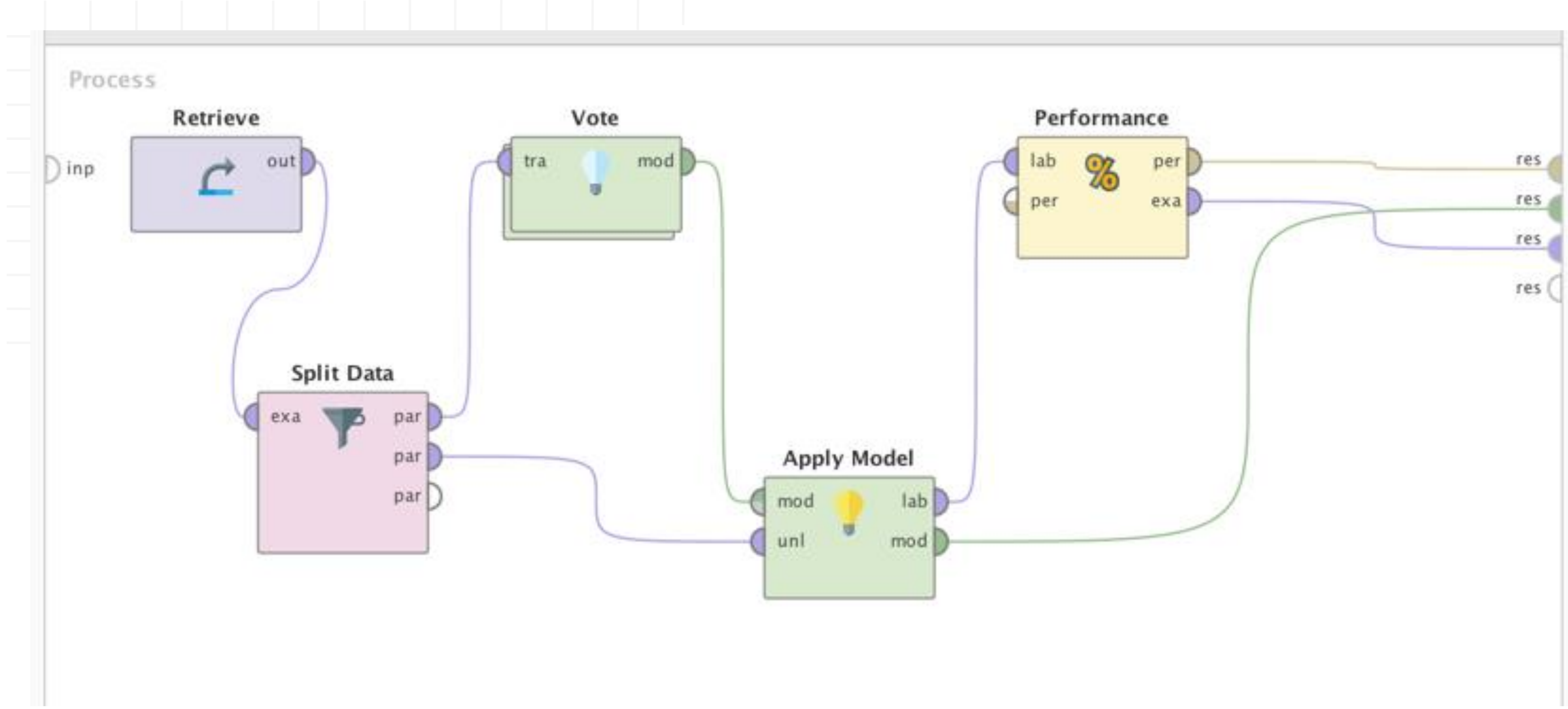
Wisdom of the Crowd

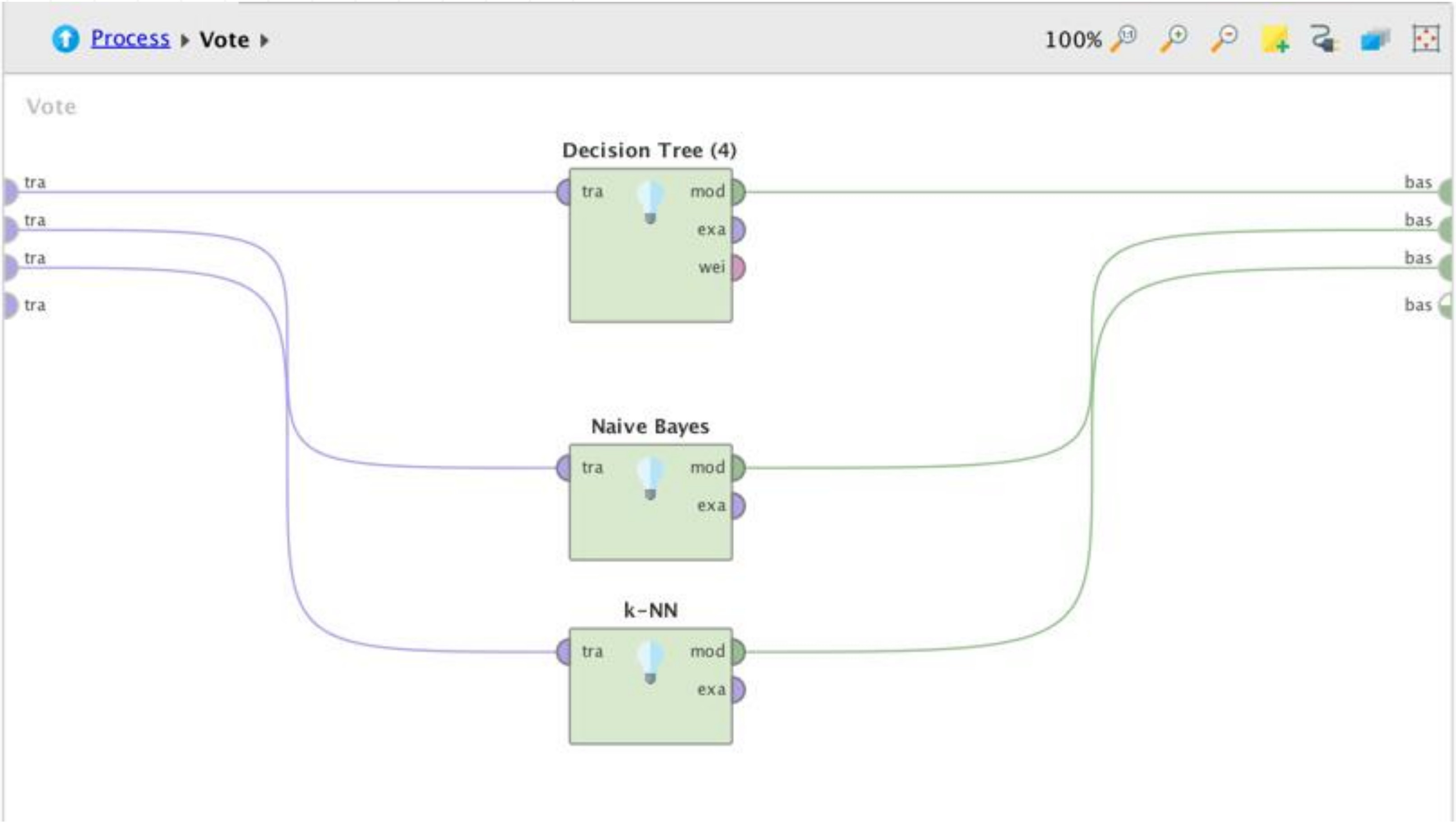
Meta learners = sum of several base models

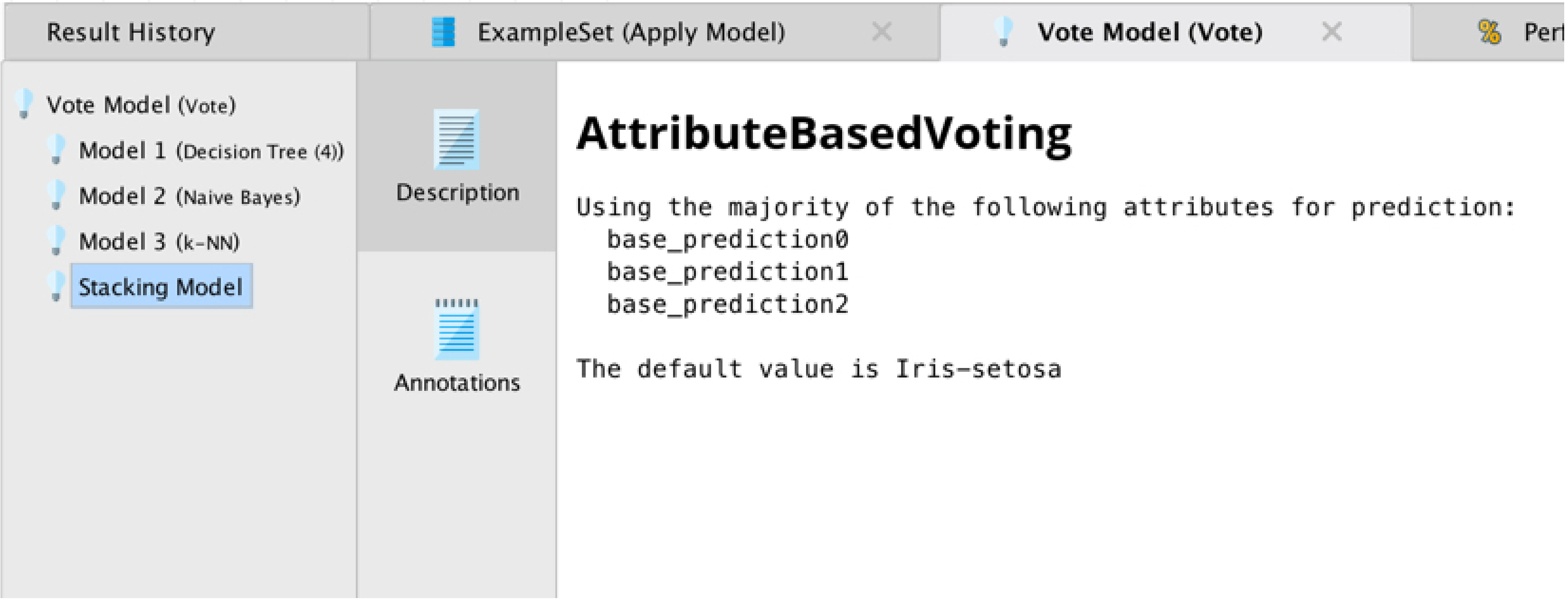
Reduces the model generalization error

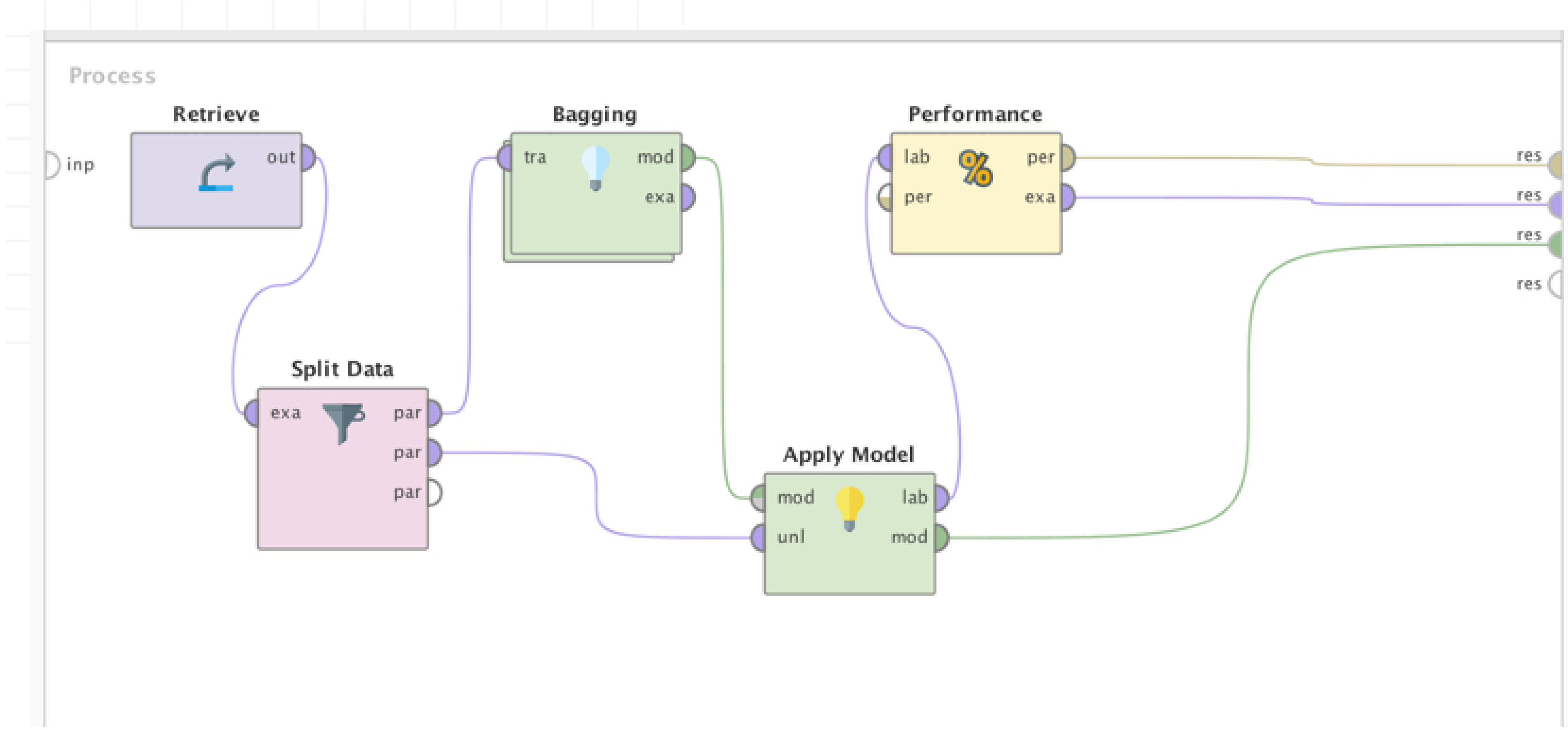
Ensemble models



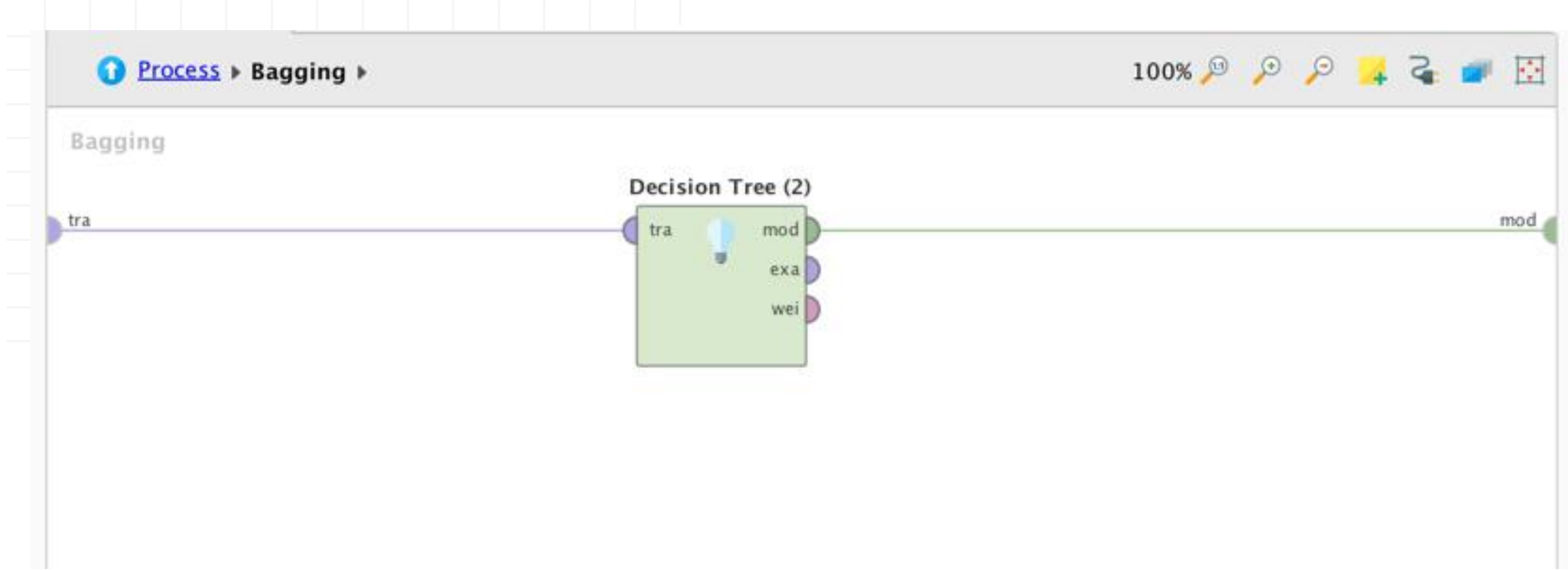

Process
Retrieve
inp
out
Vote
tra
mod
Performance
lab
per
per
exa
res
res
res
res
Split Data
exa
par
par
par
Apply Model
mod
lab
unl
mod

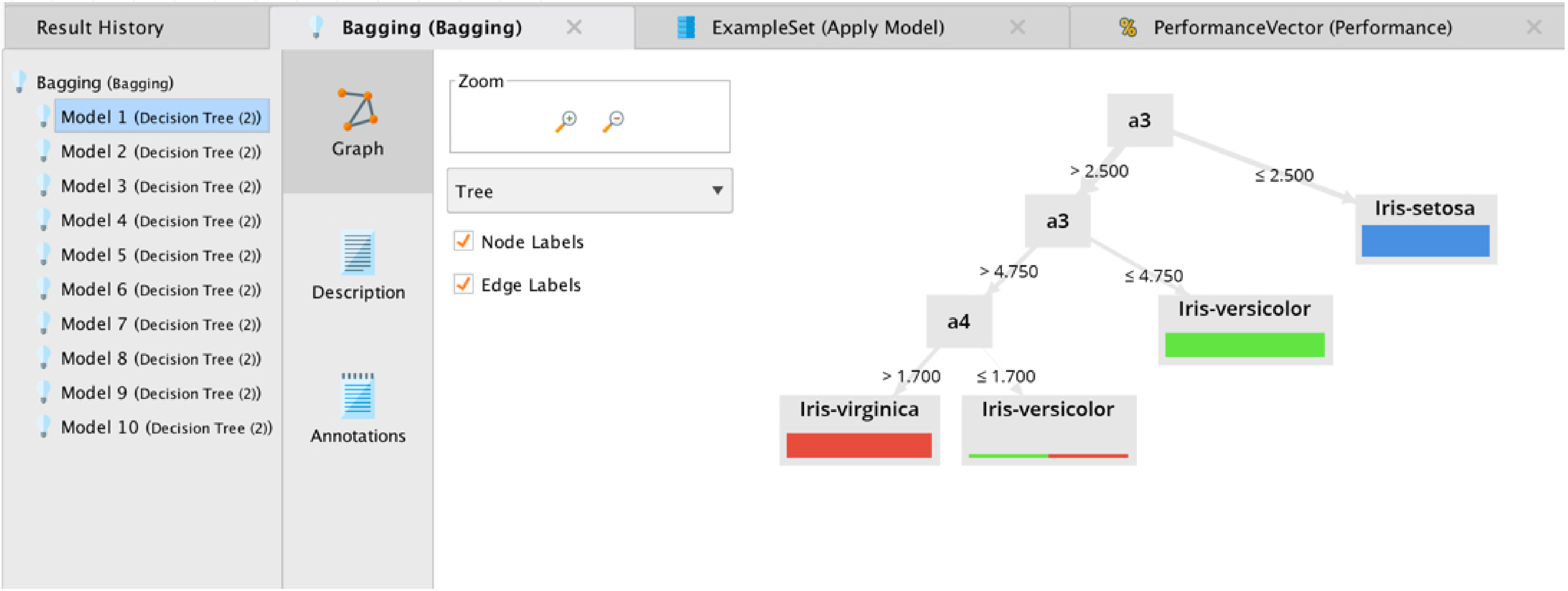

Process
Vote
100%
Vote
Decision Tree (4)
tra
mod
exa
wei
Naive Bayes
tra
mod
exa
k-NN
tra
mod
exa
tra
tra
tra
tra
bas
bas
bas
bas


Result History
ExampleSet (Apply Model)
Vote Model (Vote)
Vote Model (Vote)
Model 1 (Decision Tree (4))
Model 2 (Naive Bayes)
Model 3 (k-NN)
Stacking Model
Description
Annotations
AttributeBasedVoting
Using the majority of the following attributes for prediction:
base_prediction0
base_prediction1
base_prediction2
The default value is Iris-setosa

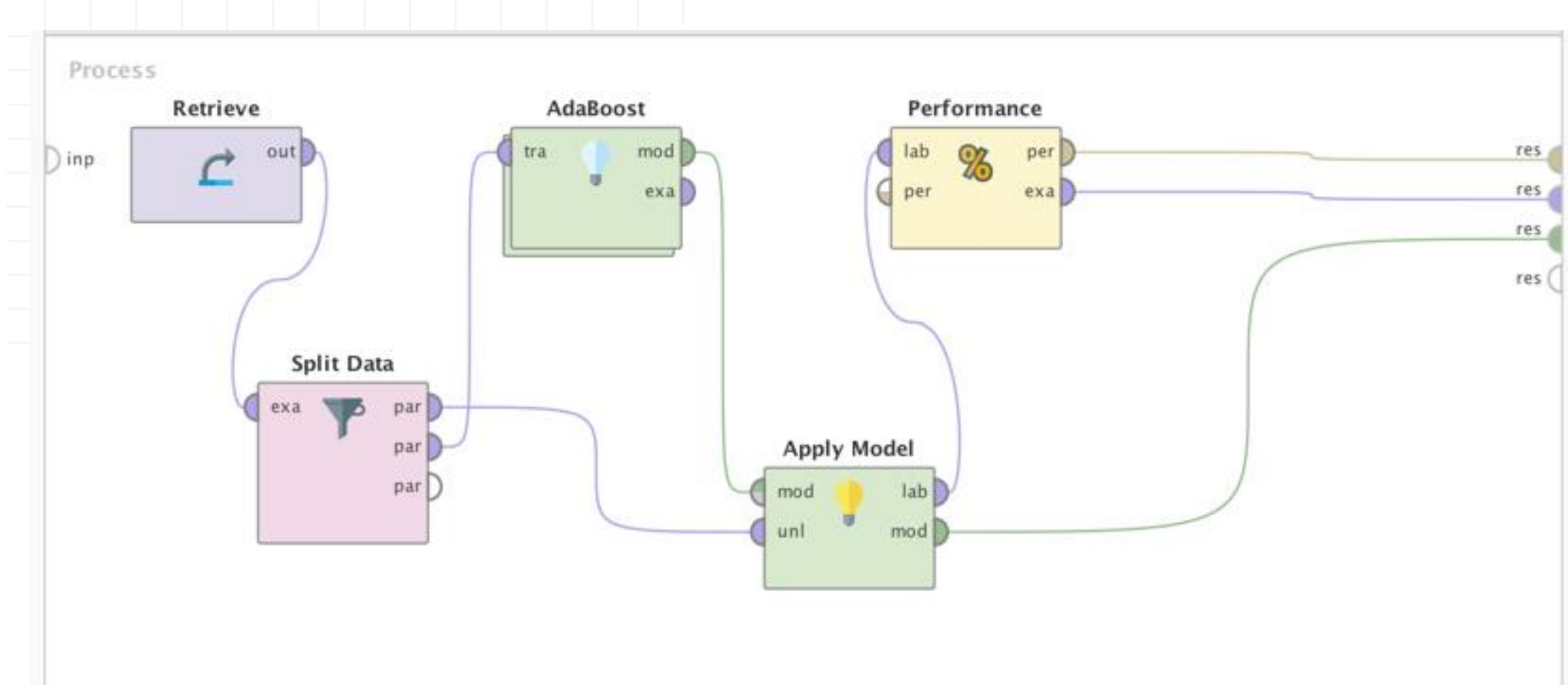


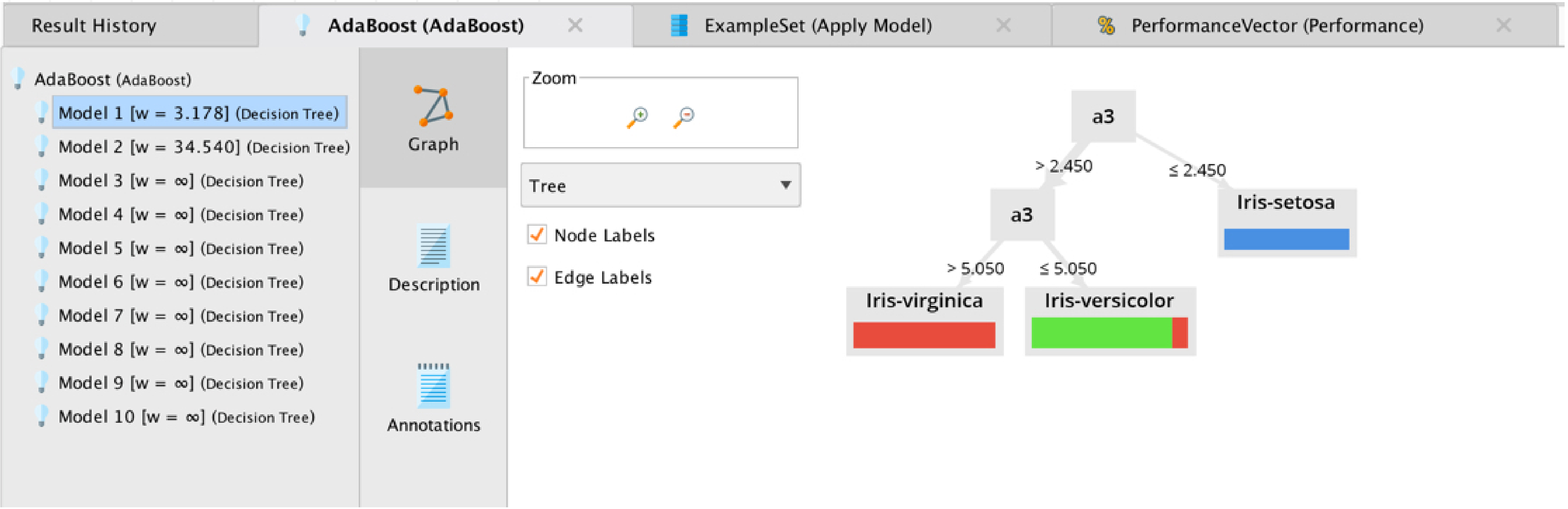
Process
Retrieve
inp
out
Bagging
tra
mod
exa
Performance
lab
per
per
exa
res
res
res
res
Split Data
exa
par
par
par
Apply Model
mod
unl
lab
mod

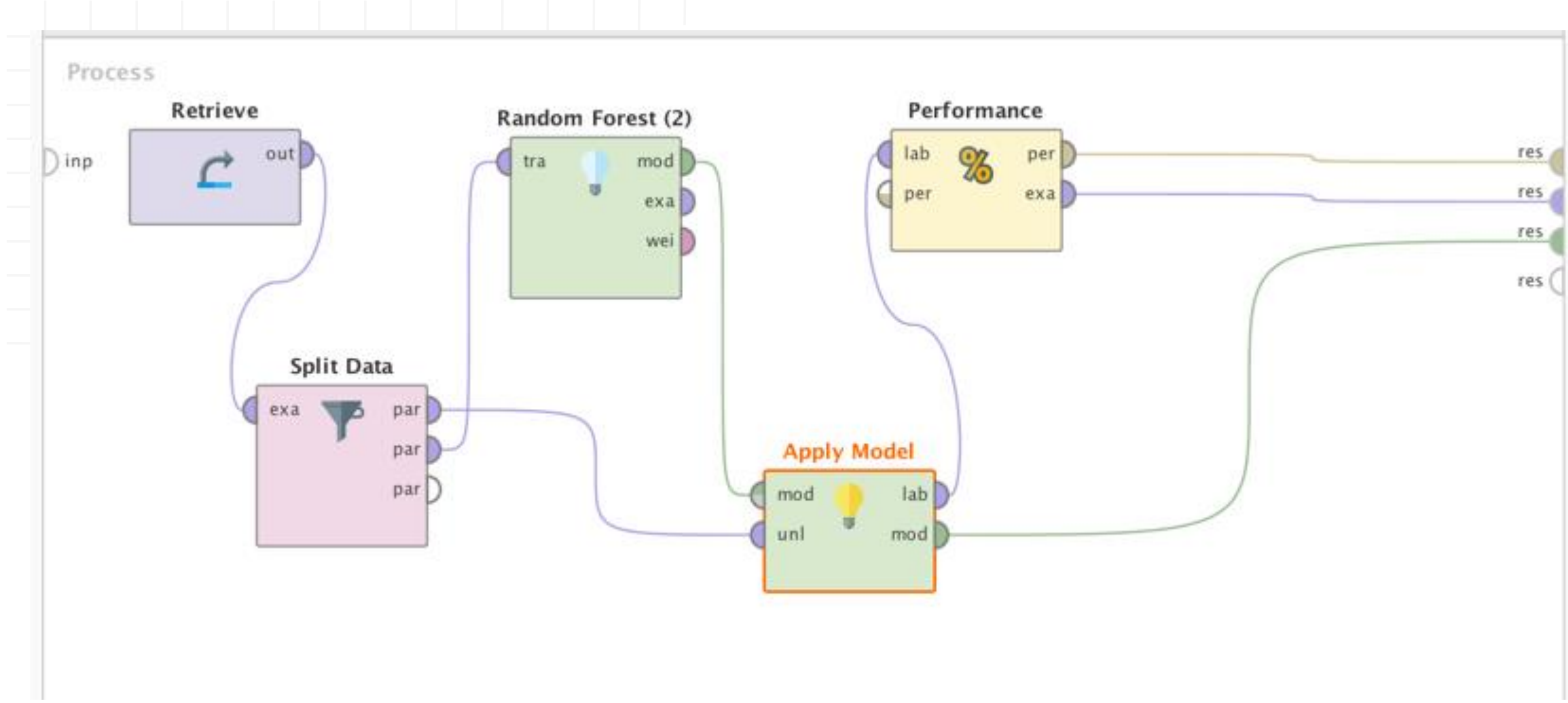

Process
Bagging
100%
Bagging
Decision Tree (2)
tra
tra
mod
exa
wei
mod



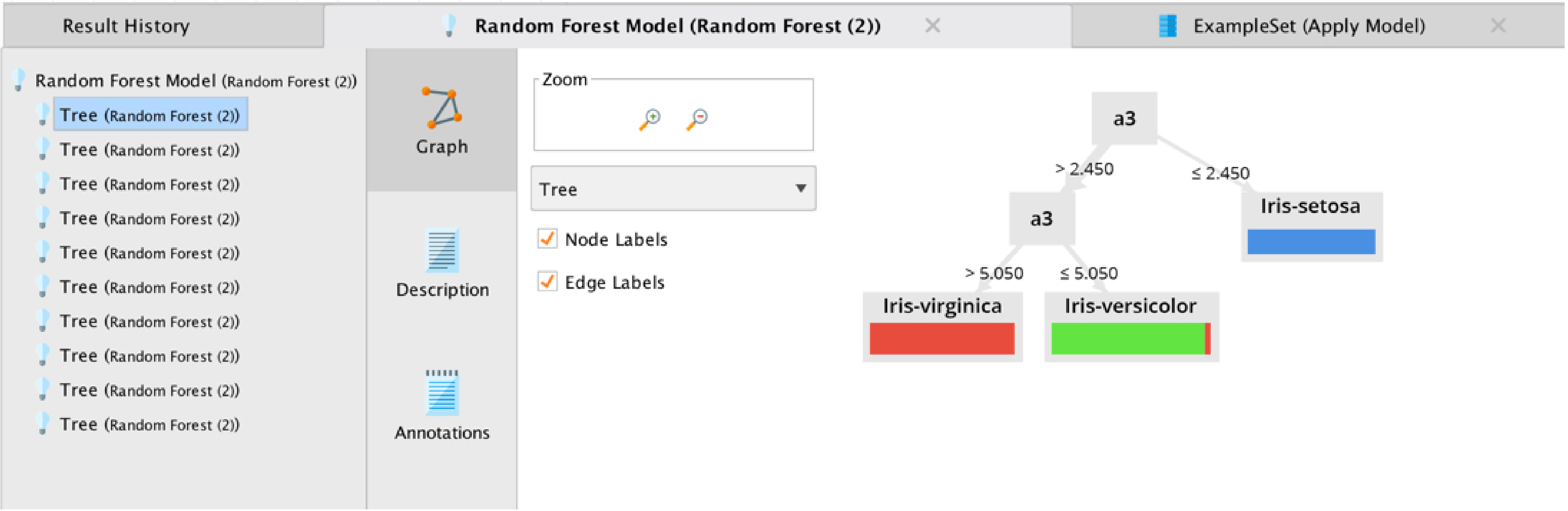
Result History
Bagging (Bagging)
ExampleSet (Apply Model)
PerformanceVector (Performance)
Bagging (Bagging)
Model 1 (Decision Tree (2))
Model 2 (Decision Tree (2))
Model 3 (Decision Tree (2))
Model 4 (Decision Tree (2))
Model 5 (Decision Tree (2))
Model 6 (Decision Tree (2))
Model 7 (Decision Tree (2))
Model 8 (Decision Tree (2))
Model 9 (Decision Tree (2))
Model 10 (Decision Tree (2))
Graph
Description
Annotations
Zoom
Tree
Node Labels
Edge Labels
a3
> 2.500
≤ 2.500
Iris-setosa
a3
> 4.750
≤ 4.750
Iris-versicolor
a4
> 1.700
≤ 1.700
Iris-virginica
Iris-versicolor


Process
Retrieve
inp
out
AdaBoost
tra
mod
exa
Performance
lab
per
per
exa
res
res
res
res
Split Data
exa
par
par
par
Apply Model
mod
unl
lab
mod


Result History
AdaBoost (AdaBoost)
ExampleSet (Apply Model)
PerformanceVector (Performance)
AdaBoost (AdaBoost)
Model 1 [w = 3.178] (Decision Tree)
Model 2 [w = 34.540] (Decision Tree)
Model 3 [w = ∞] (Decision Tree)
Model 4 [w = ∞] (Decision Tree)
Model 5 [w = ∞] (Decision Tree)
Model 6 [w = ∞] (Decision Tree)
Model 7 [w = ∞] (Decision Tree)
Model 8 [w = ∞] (Decision Tree)
Model 9 [w = ∞] (Decision Tree)
Model 10 [w = ∞] (Decision Tree)
Graph
Description
Annotations
Zoom
Tree
Node Labels
Edge Labels
a3
> 2.450
≤ 2.450
a3
Iris-setosa
> 5.050
≤ 5.050
Iris-virginica
Iris-versicolor



Process
Retrieve
inp
out
Random Forest (2)
tra
mod
exa
wei
Performance
lab
per
per
exa
res
res
res
res
Split Data
exa
par
par
par
Apply Model
mod
unl
lab
mod



Result History
Random Forest Model (Random Forest (2))
ExampleSet (Apply Model)
Random Forest Model (Random Forest (2))
Tree (Random Forest (2))
Tree (Random Forest (2))
Tree (Random Forest (2))
Tree (Random Forest (2))
Tree (Random Forest (2))
Tree (Random Forest (2))
Tree (Random Forest (2))
Tree (Random Forest (2))
Tree (Random Forest (2))
Tree (Random Forest (2))
Graph
Description
Annotations
Zoom
Tree
Node Labels
Edge Labels
a3
> 2.450
≤ 2.450
a3
Iris-setosa
> 5.050
≤ 5.050
Iris-virginica
Iris-versicolor