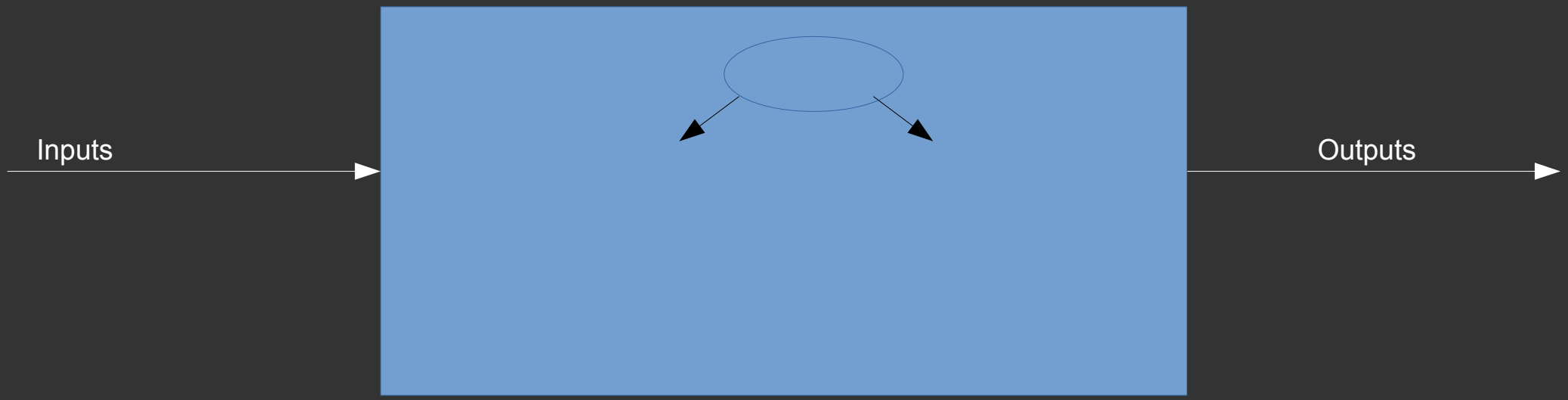
Decision Tree

Supervised Learning,
Information Theory,
Scalable,
Human Comprehensible

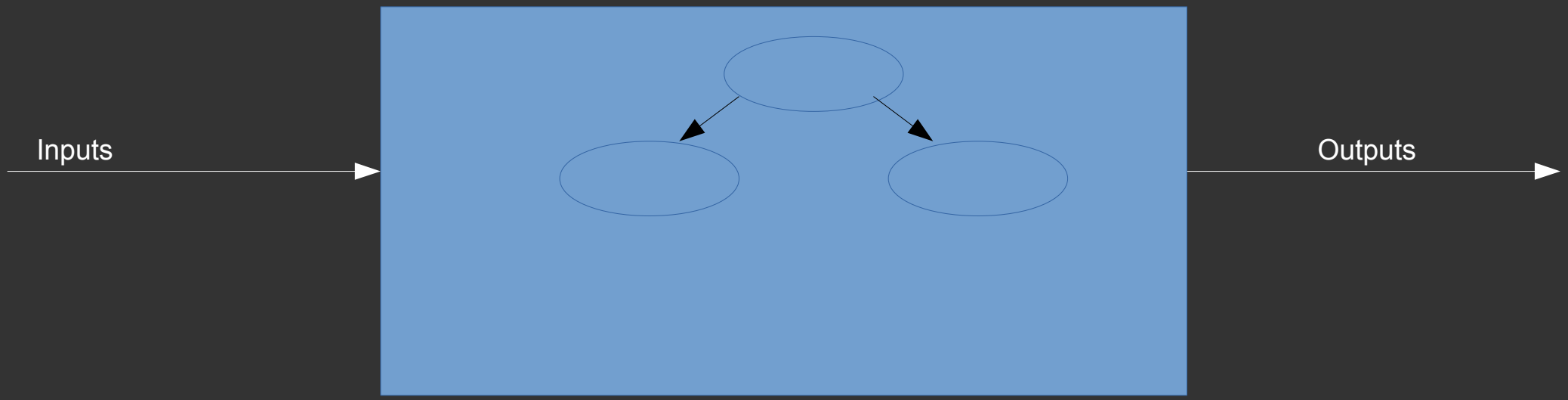
Allen Grimm
3/16/2016
allen@GrimmScience.com
@GrimmScientist

Supervised Learning (Black Box)

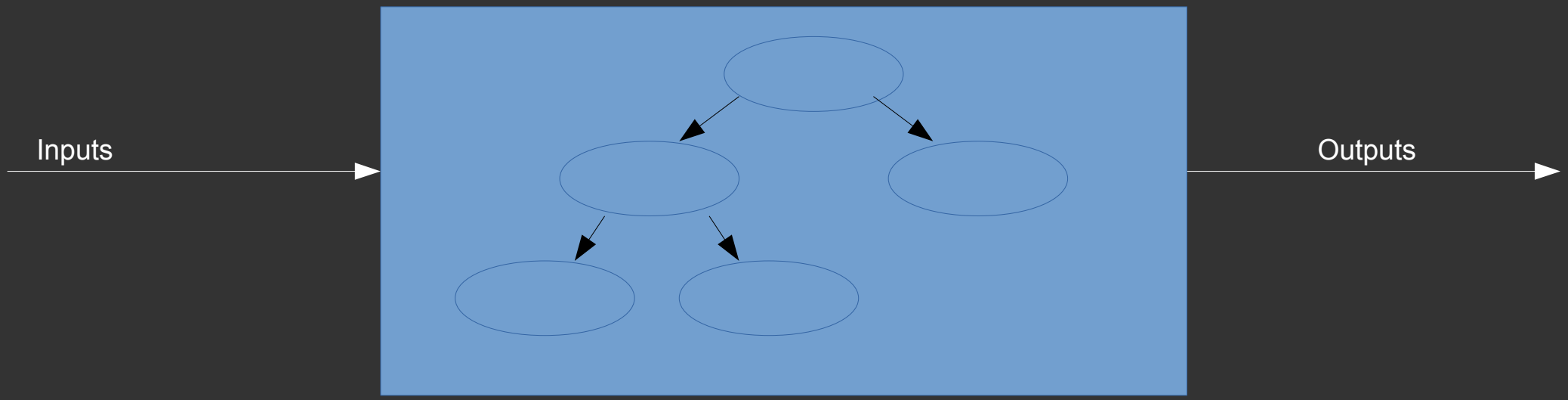
Black Box



Black Box



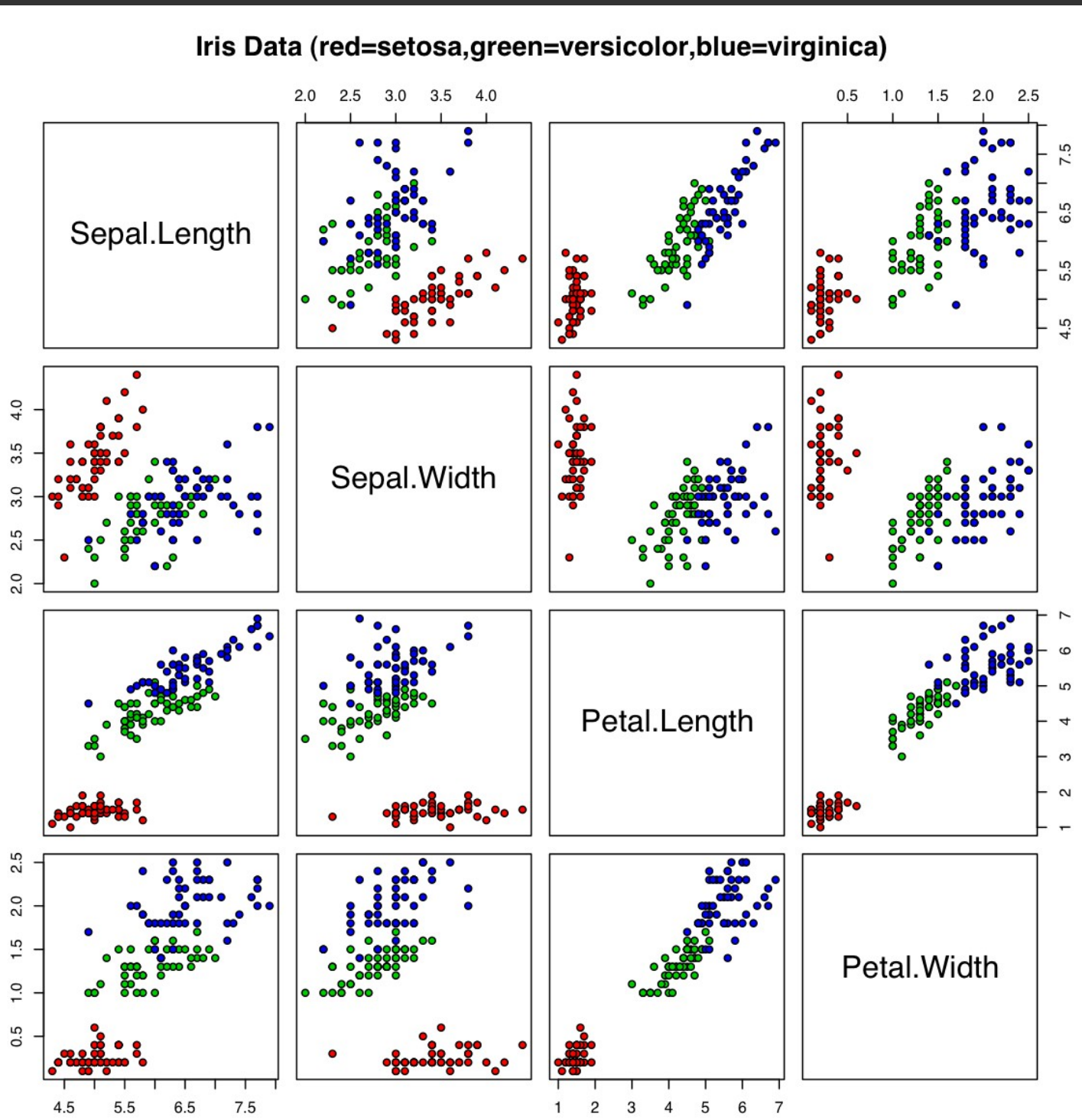
Black Box



Introducing: Fisher's Iris Data

- shape of data: (150, 4)
- iv names: ['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)', 'petal width (cm)']
- shape of labels: (150,)
- label names: ['setosa' 'versicolor' 'virginica']

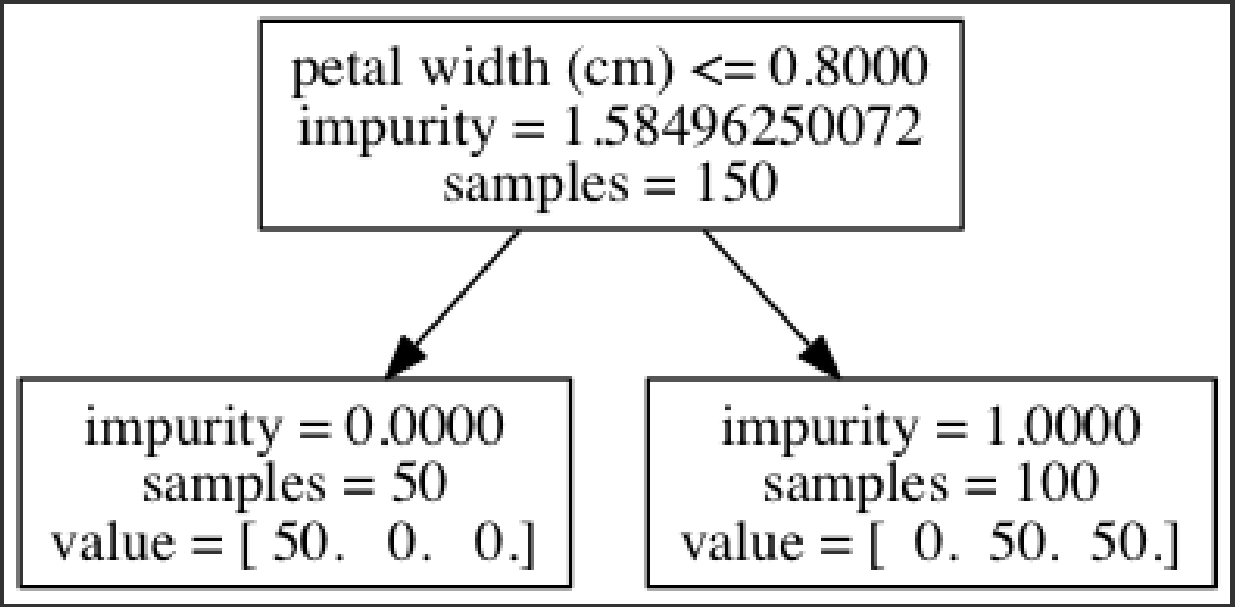
- Picking 15 samples, with labels:
  - [5.1 3.5 1.4 0.2] 0
  - [5.7 4.4 1.5 0.4] 0
  - [4.8 3.1 1.6 0.2] 0
  - [4.8 3.   1.4 0.3] 0
  - [5.   2.   3.5 1.] 1
  - [6.6 3.   4.4 1.4] 1
  - [5.5 2.6 4.4 1.2] 1
  - [7.6 3.   6.6 2.1] 2
  - [6.9 3.2 5.7 2.3] 2
  - [7.7 3.   6.1 2.3] 2


Iris Data (red=setosa,green=versicolor,blue=virginica)
Sepal.Length
Sepal.Width
Petal.Length
Petal.Width

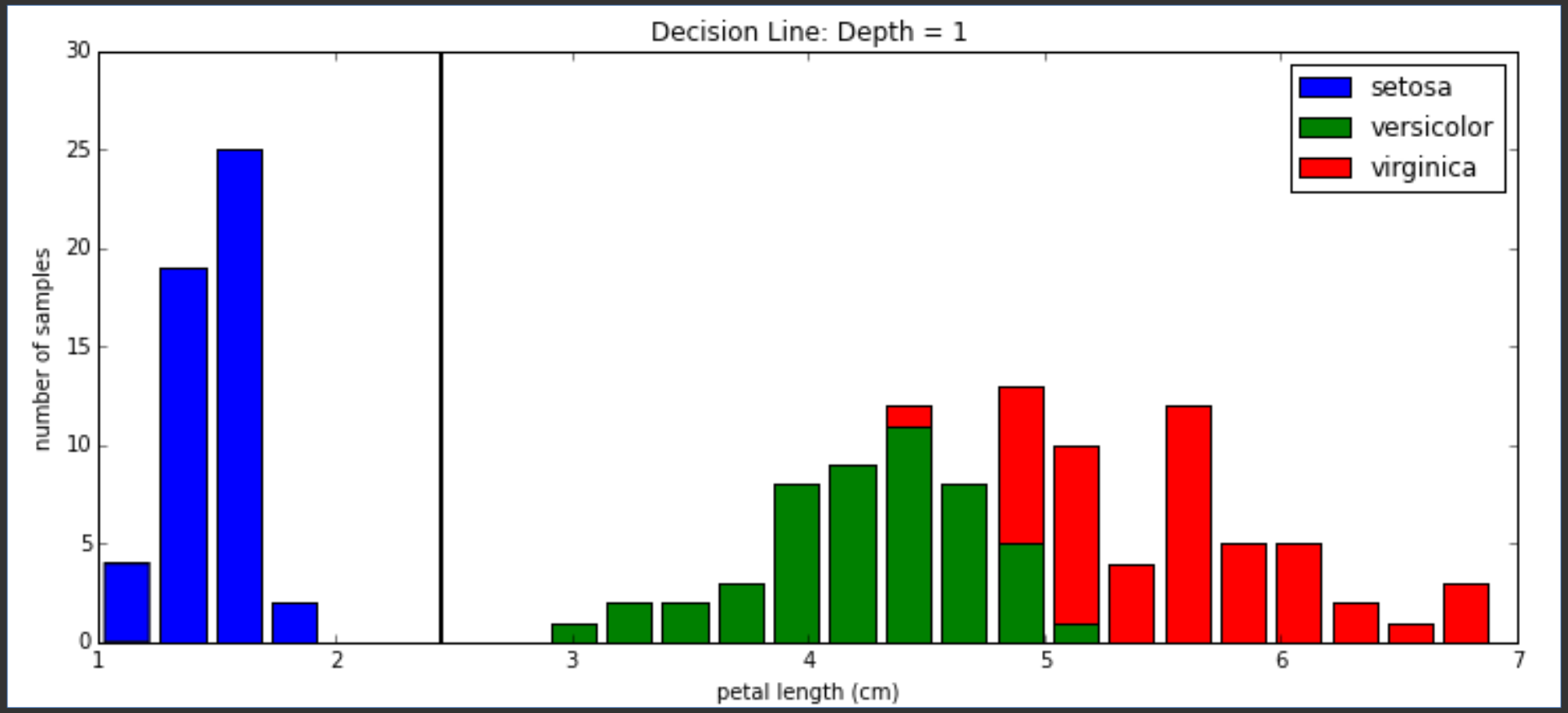
Petal Length

!!!!! (add graph without line)

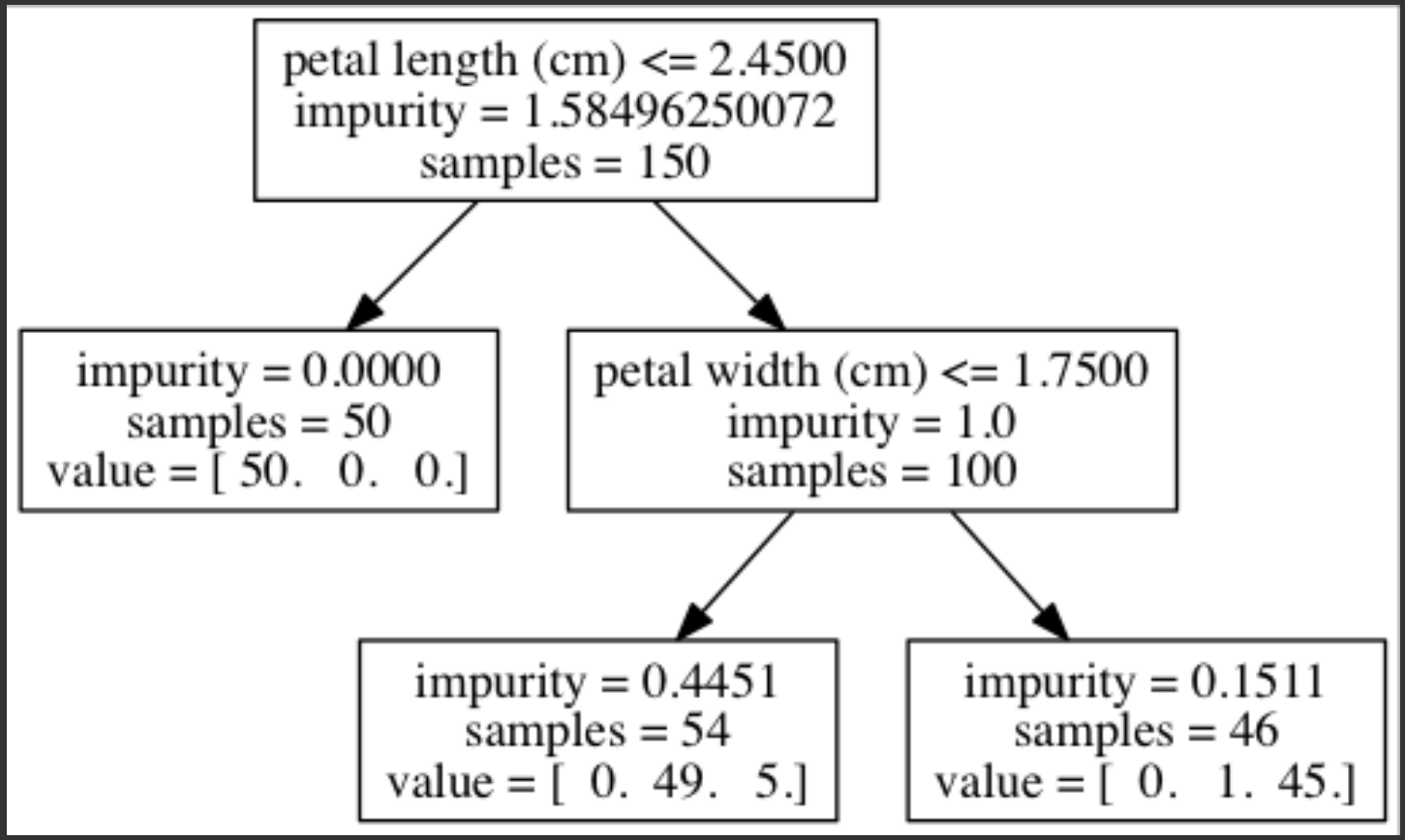
Depth = 1


petal width (cm) <= 0.8000
impurity = 1.58496250072
samples = 150
impurity = 0.0000
samples = 50
value = [50. 0. 0.]
impurity = 1.0000
samples = 100
value = [0. 50. 50.]

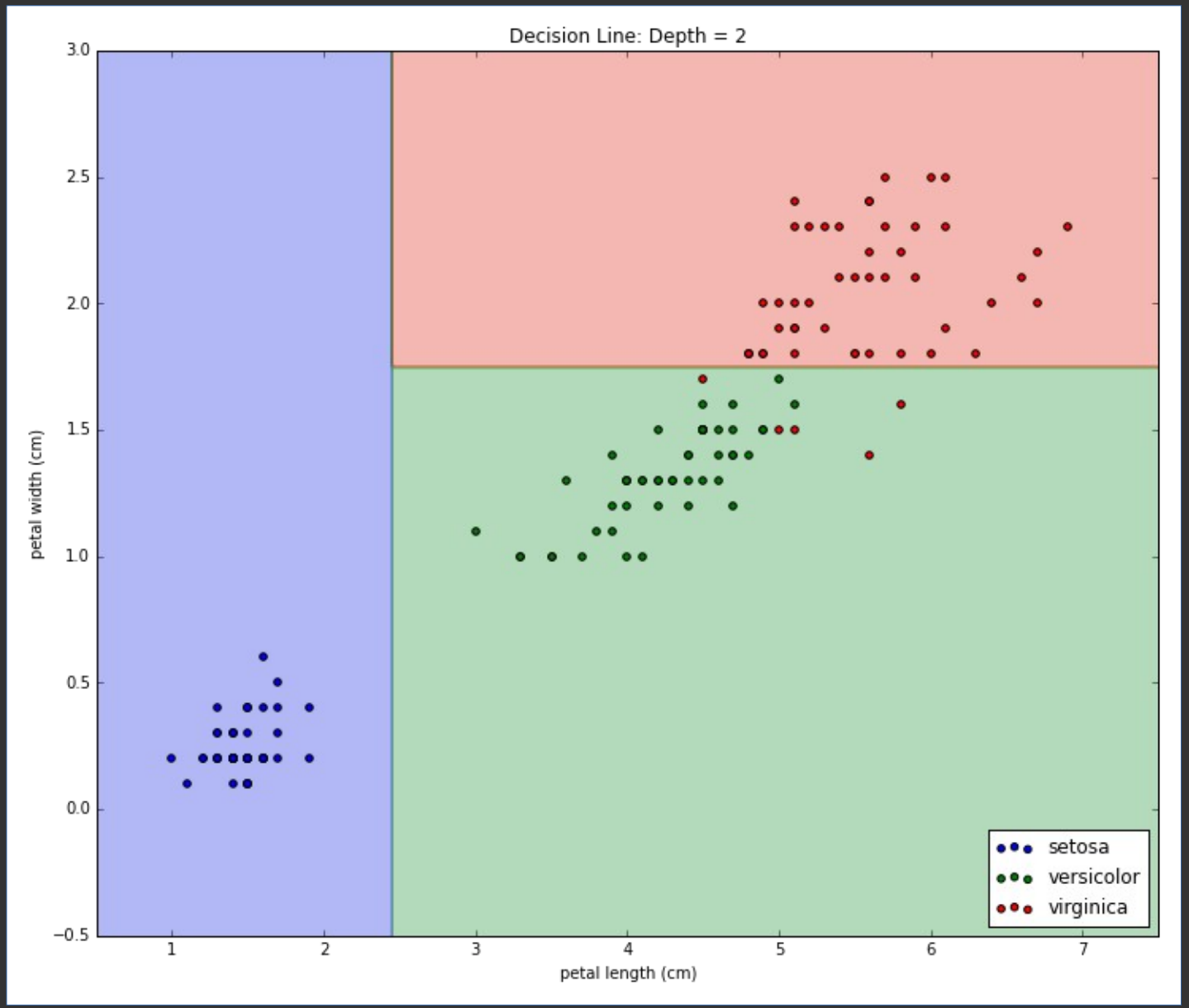
Petal Length, with decision line



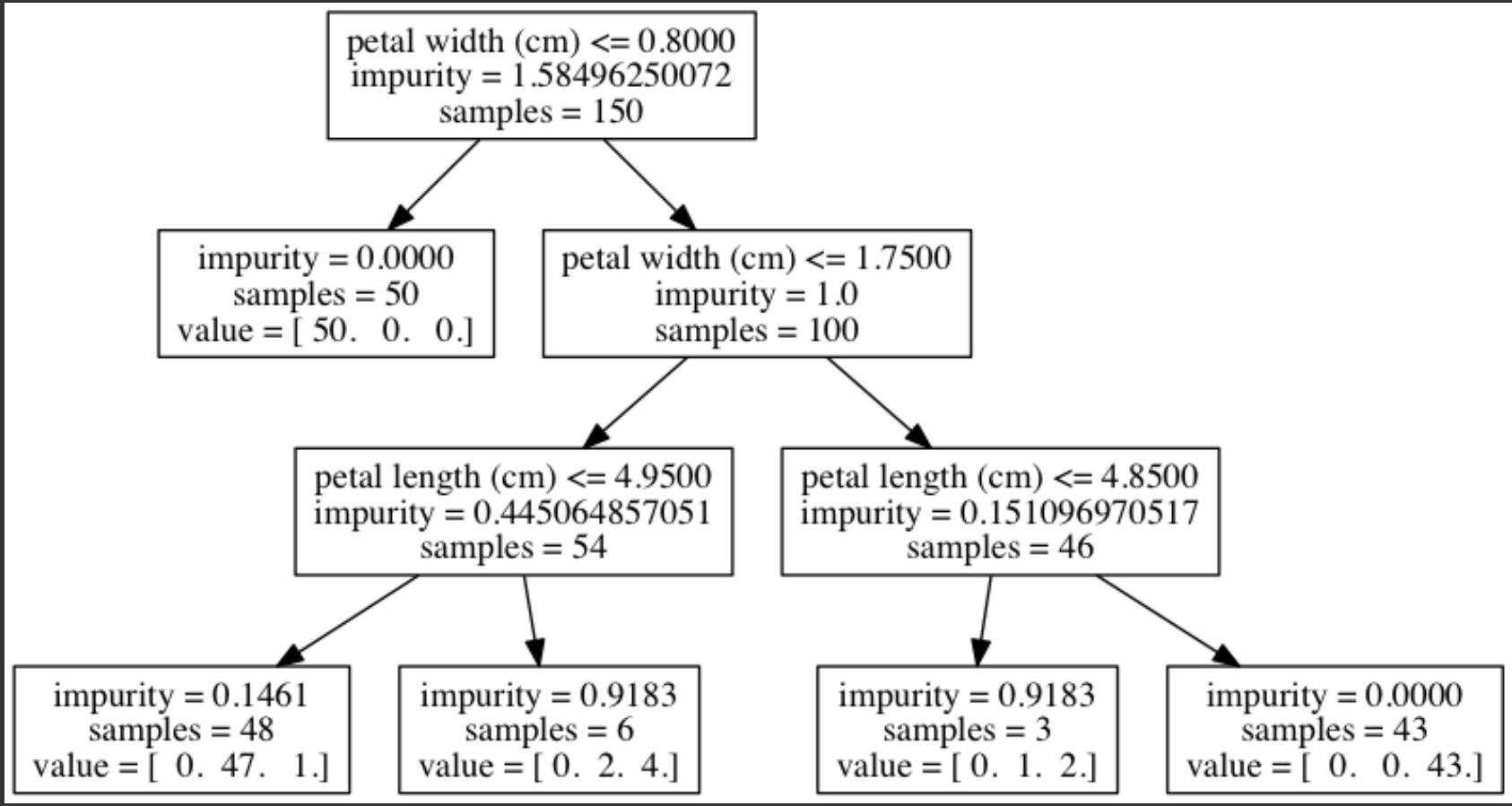
Depth = 2


petal length (cm) <= 2.4500
impurity = 1.58496250072
samples = 150
impurity = 0.0000
samples = 50
value = [50. 0. 0.]
petal width (cm) <= 1.7500
impurity = 1.0
samples = 100
impurity = 0.4451
samples = 54
value = [0. 49. 5.]
impurity = 0.1511
samples = 46
value = [0. 1. 45.]

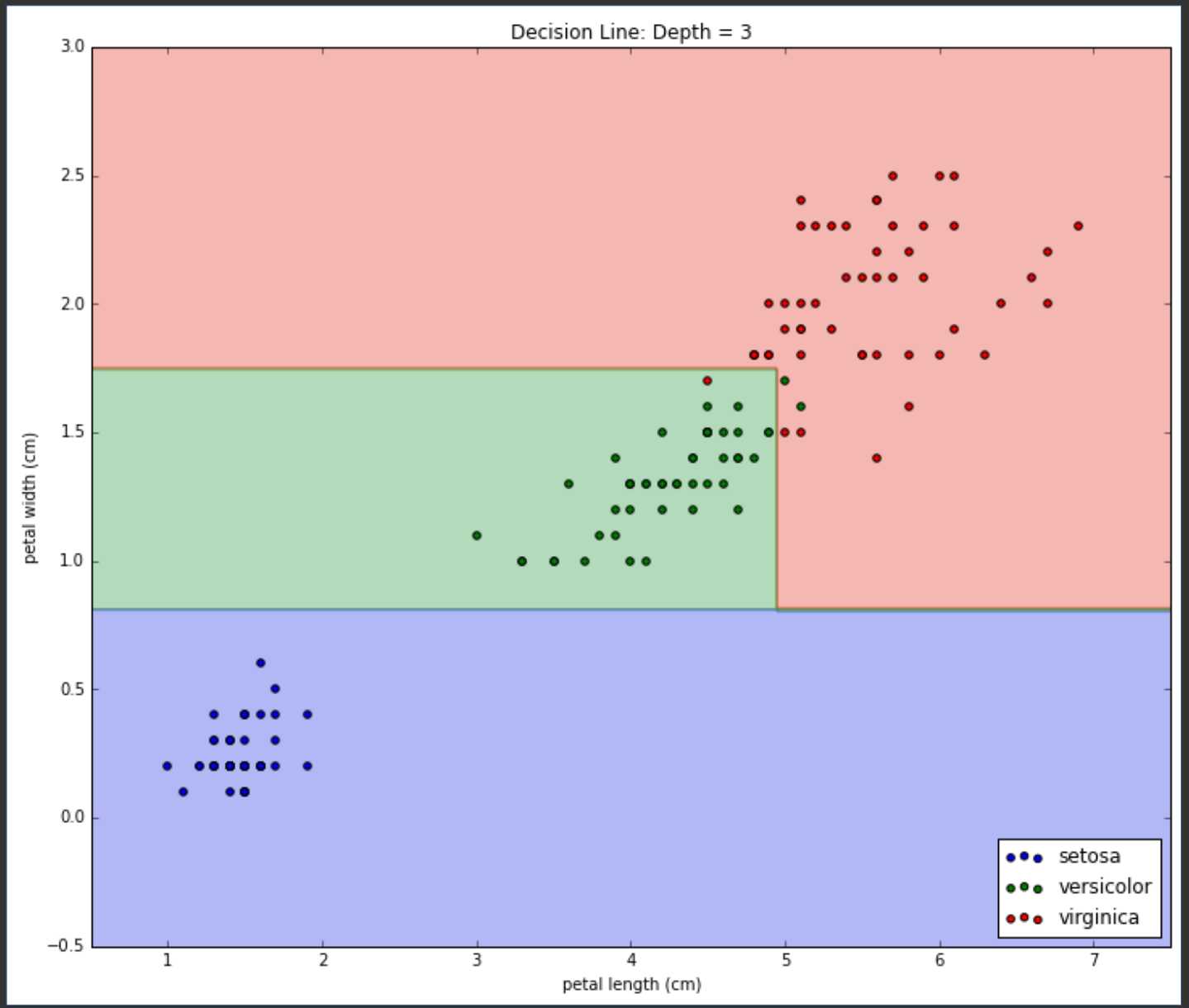
Petal Length and Petal Width. d=2



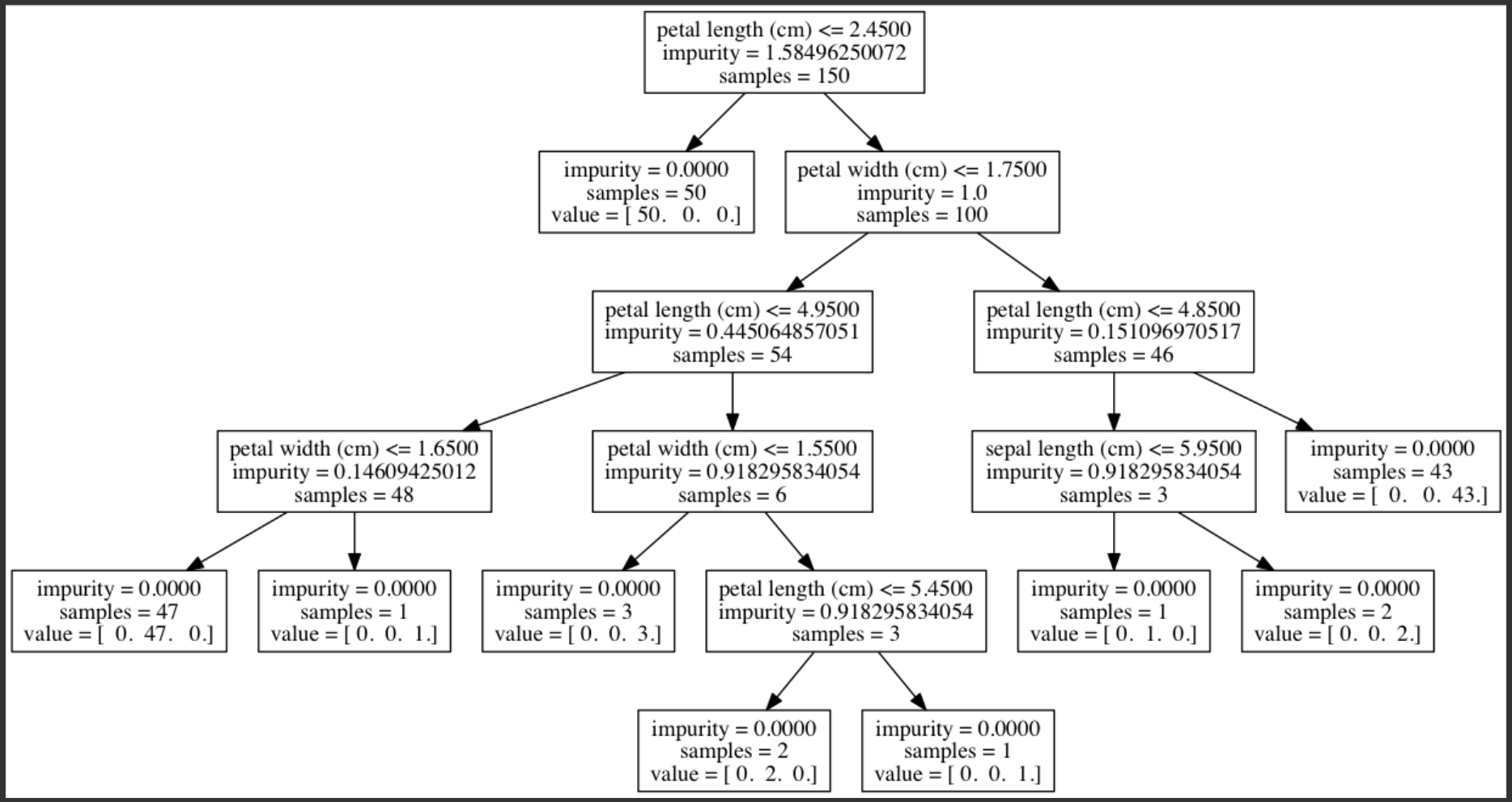
Depth = 3


petal width (cm) <= 0.8000
impurity = 1.58496250072
samples = 150
impurity = 0.0000
samples = 50
value = [50. 0. 0.]
petal width (cm) <= 1.7500
impurity = 1.0
samples = 100
petal length (cm) <= 4.9500
impurity = 0.445064857051
samples = 54
petal length (cm) <= 4.8500
impurity = 0.151096970517
samples = 46
impurity = 0.1461
samples = 48
value = [0. 47. 1.]
impurity = 0.9183
samples = 6
value = [0. 2. 4.]
impurity = 0.9183
samples = 3
value = [0. 1. 2.]
impurity = 0.0000
samples = 43
value = [0. 0. 43.]

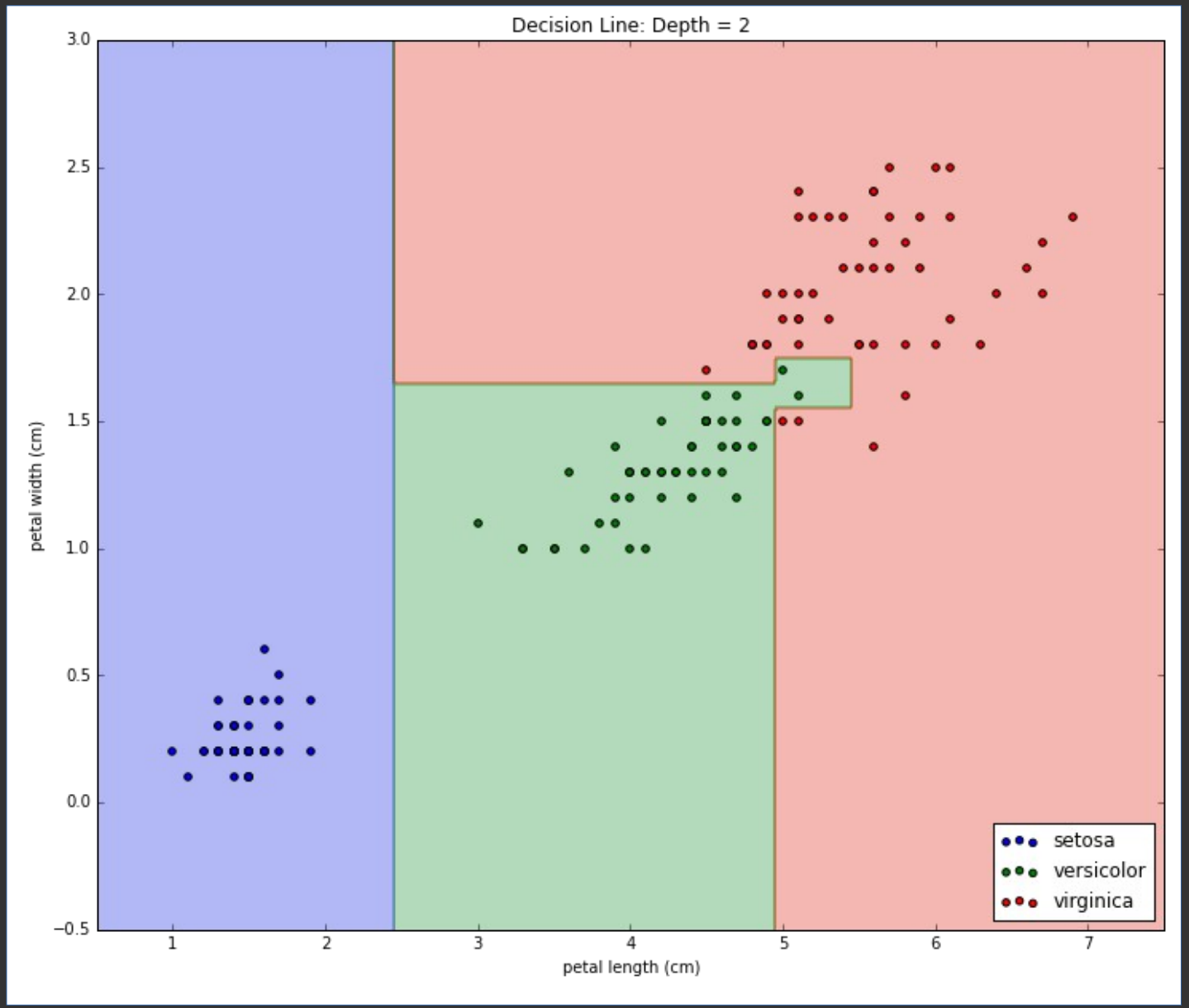
Petal Length and Petal Width. d=3



Depth = *


petal length (cm) <= 2.4500
impurity = 1.58496250072
samples = 150
impurity = 0.0000
samples = 50
value = [50. 0. 0.]
petal width (cm) <= 1.7500
impurity = 1.0
samples = 100
petal length (cm) <= 4.9500
impurity = 0.445064857051
samples = 54
petal length (cm) <= 4.8500
impurity = 0.151096970517
samples = 46
petal width (cm) <= 1.6500
impurity = 0.14609425012
samples = 48
petal width (cm) <= 1.5500
impurity = 0.918295834054
samples = 6
sepal length (cm) <= 5.9500
impurity = 0.918295834054
samples = 3
impurity = 0.0000
samples = 43
value = [0. 0. 43.]
impurity = 0.0000
samples = 47
value = [0. 47. 0.]
impurity = 0.0000
samples = 1
value = [0. 0. 1.]
impurity = 0.0000
samples = 3
value = [0. 0. 3.]
petal length (cm) <= 5.4500
impurity = 0.918295834054
samples = 3
impurity = 0.0000
samples = 1
value = [0. 1. 0.]
impurity = 0.0000
samples = 2
value = [0. 0. 2.]
impurity = 0.0000
samples = 2
value = [0. 2. 0.]
impurity = 0.0000
samples = 1
value = [0. 0. 1.]

Petal Length and Petal Width. d=*



Entropy

- To the notebook!