



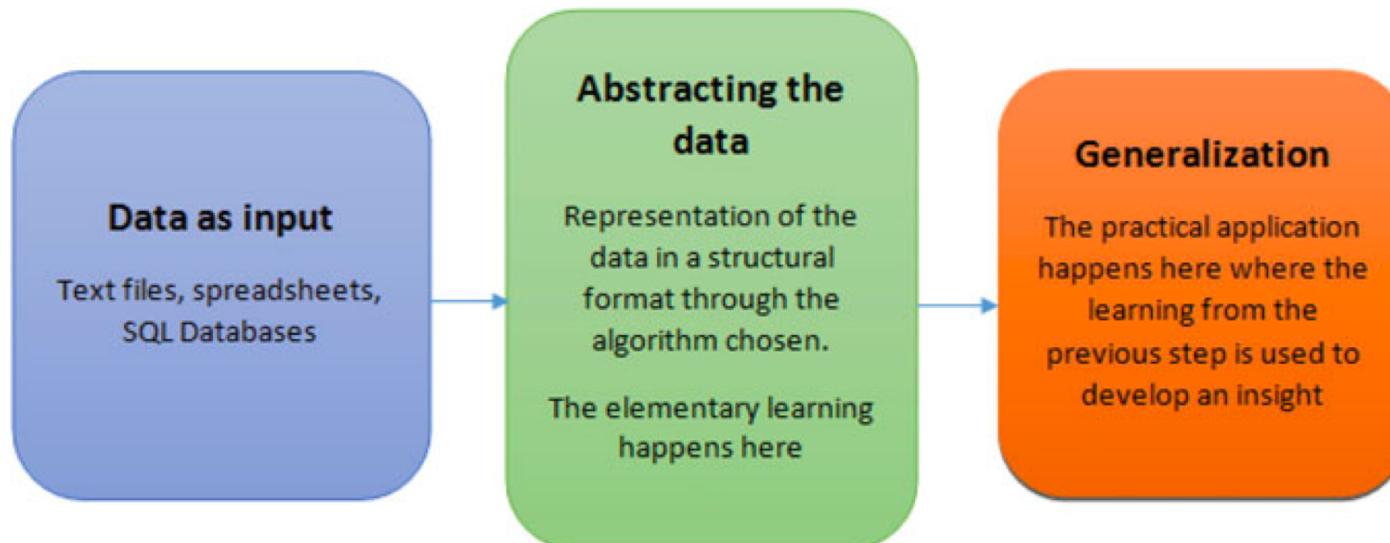
MACHINE  
LEARNING

Logistic Regression

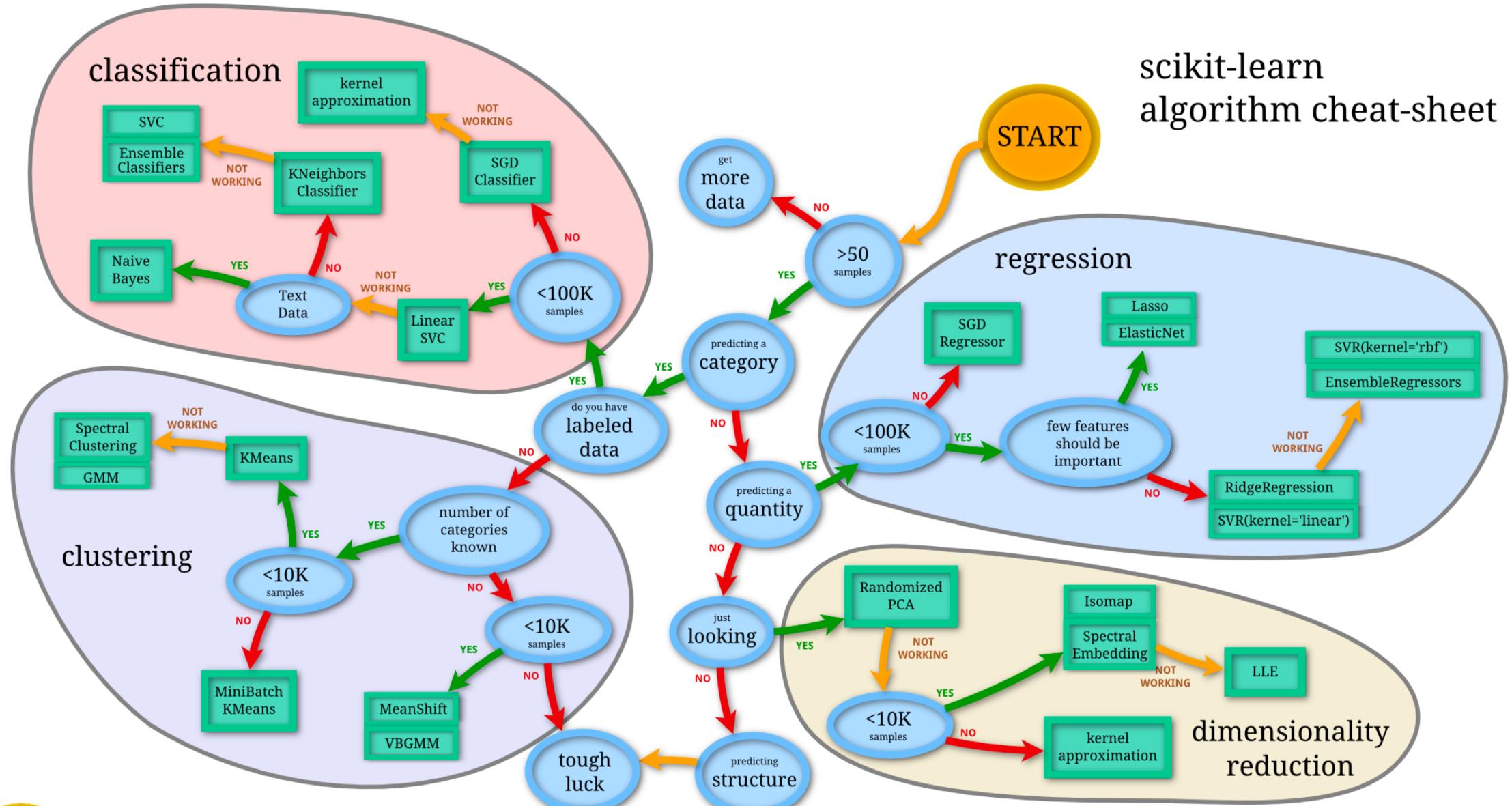
Oct 25<sup>th</sup> 2021

# Machine Learning

- Applications
- “*Learn without being explicitly programmed*”

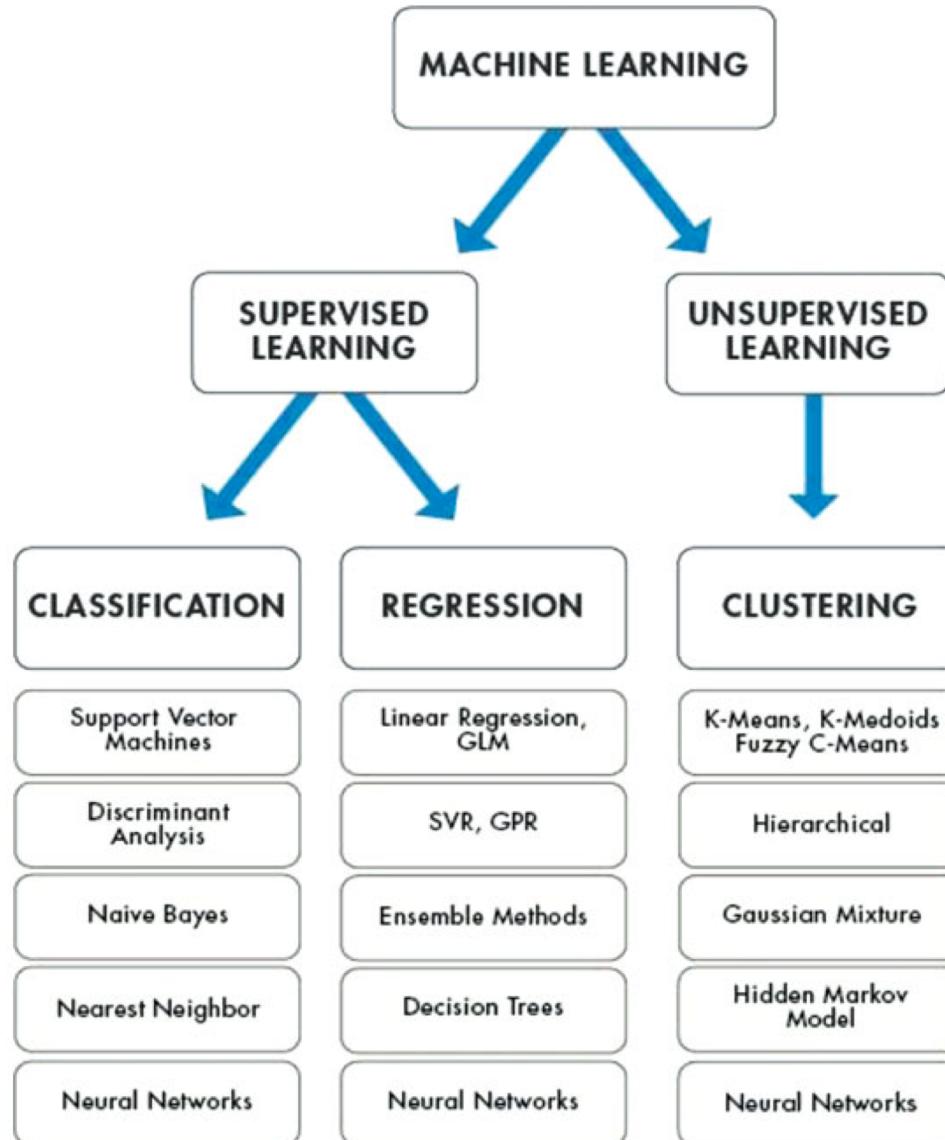


# scikit-learn algorithm cheat-sheet

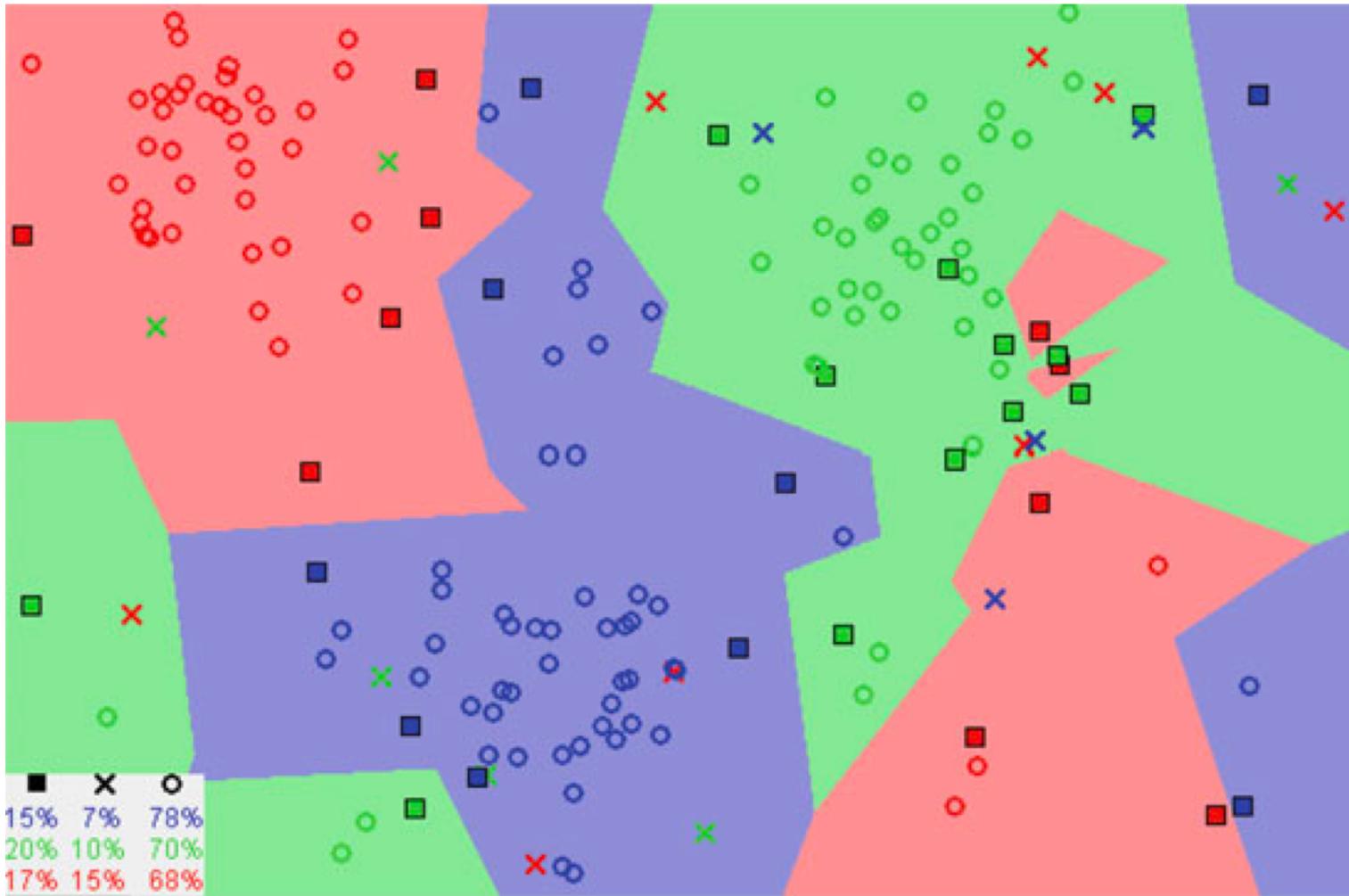


*Back*

scikit  
learn



# K-Nearest Neighbor



# Logistic Regression

# Odds

Table 4-1. Probabilities and the corresponding odds.

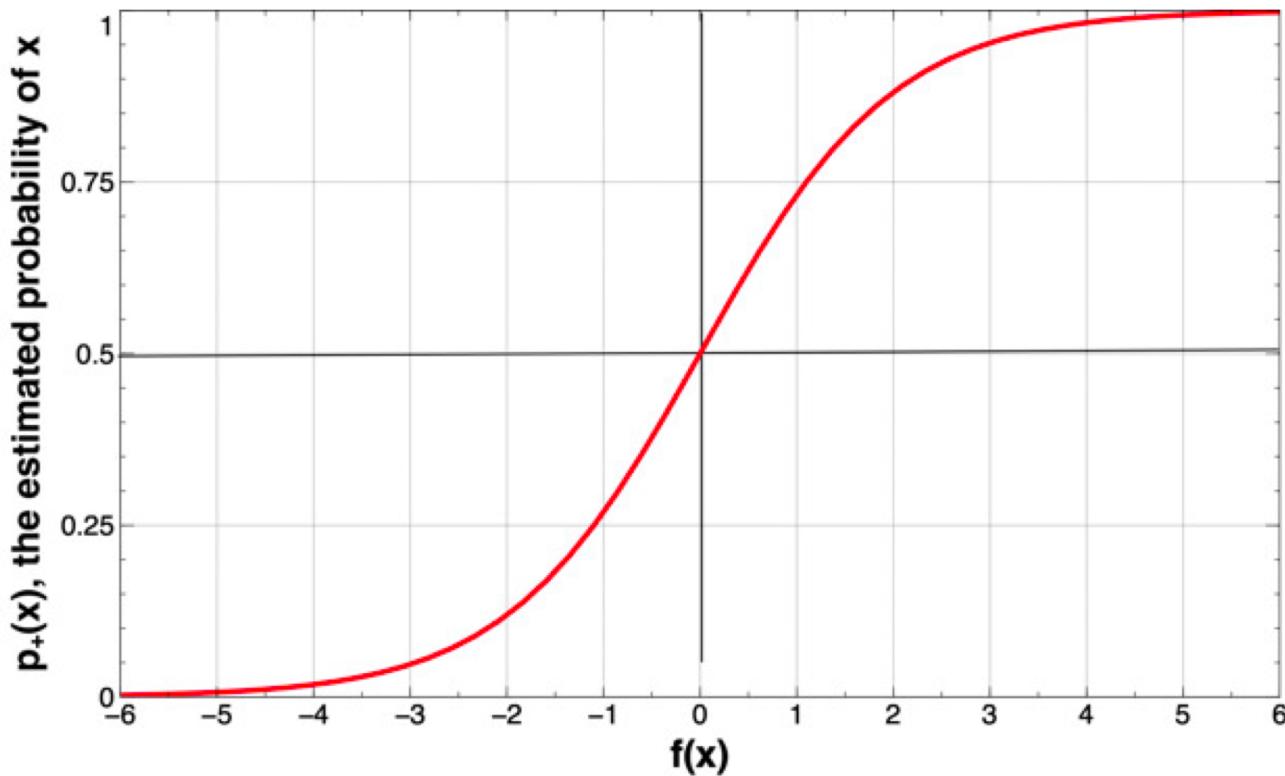
Probability	Corresponding odds
0.5	50:50 or 1
0.9	90:10 or 9
0.999	999:1 or 999
0.01	1:99 or 0.0101
0.001	1:999 or 0.001001

Table 4-2. Probabilities, odds, and the corresponding log-odds.

Probability	Odds	Log-odds
0.5	50:50 or 1	0
0.9	90:10 or 9	2.19
0.999	999:1 or 999	6.9
0.01	1:99 or 0.0101	-4.6
0.001	1:999 or 0.001001	-6.9

$$\log \left( \frac{p_+(\mathbf{x})}{1 - p_+(\mathbf{x})} \right) = f(\mathbf{x}) = w_0 + w_1 x_1 + w_2 x_2 + \dots$$

# Sigmoid Curve



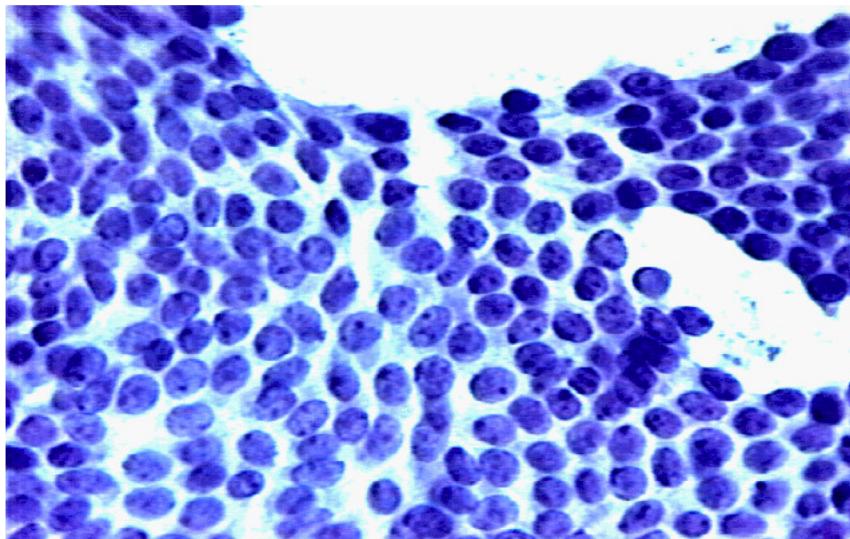
$$p_+(x) = \frac{1}{1 + e^{-f(x)}}$$

# Application of Logistic Regression

- The Wisconsin Breast Cancer Dataset



# Wisconsin Breast Cancer dataset



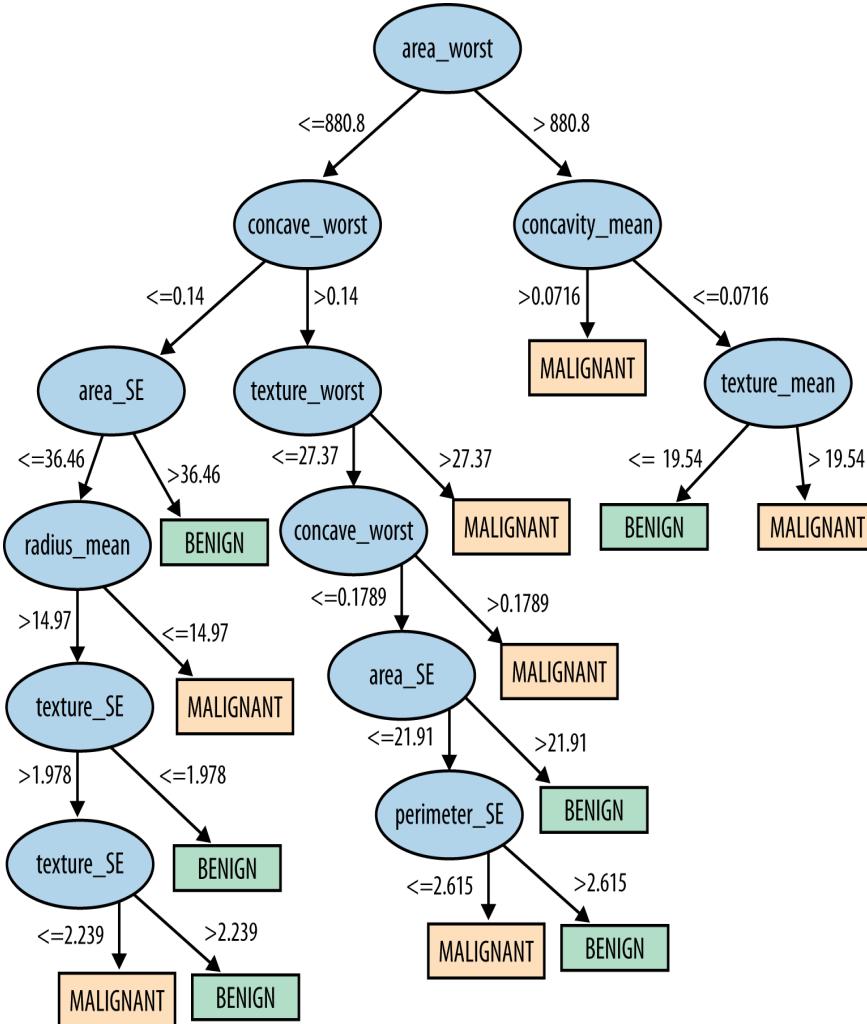
Attribute name	Description
RADIUS	<i>Mean of distances from center to points on the perimeter</i>
TEXTURE	<i>Standard deviation of grayscale values</i>
PERIMETER	<i>Perimeter of the mass</i>
AREA	<i>Area of the mass</i>
SMOOTHNESS	<i>Local variation in radius lengths</i>
COMPACTNESS	<i>Computed as: perimeter<sup>2</sup>/area – 1.0</i>
CONCAVITY	<i>Severity of concave portions of the contour</i>
CONCAVE POINTS	<i>Number of concave portions of the contour</i>
SYMMETRY	<i>A measure of the symmetry of the nucleii</i>
FRACTAL DIMENSION	<i>'Coastline approximation' – 1.0</i>
DIAGNOSIS (Target)	<i>Diagnosis of cell sample: malignant or benign</i>

- From each of these basic characteristics, three values were computed: the mean (\_mean), standard error (\_SE), and “worst” or largest

# Logistic Regression Learnt

Attribute	Weight (learned parameter)
SMOOTHNESS_worst	22.3
CONCAVE_mean	19.47
CONCAVE_worst	11.68
SYMMETRY_worst	4.99
CONCAVITY_worst	2.86
CONCAVITY_mean	2.34
RADIUS_worst	0.25
TEXTURE_worst	0.13
AREA_SE	0.06
TEXTURE_mean	0.03
TEXTURE_SE	-0.29
COMPACTNESS_mean	-7.1
COMPACTNESS_SE	-27.87
$w_0$ (intercept)	-17.7

# Decision Tree Learnt



# Accuracy

- Logistic Regression : 98.9 %
- Decision Tree (J48) : 99.1 %

# Non-linear Functions

- Linear functions can actually represent nonlinear models, if we include more complex features in the functions

