AI Seminar

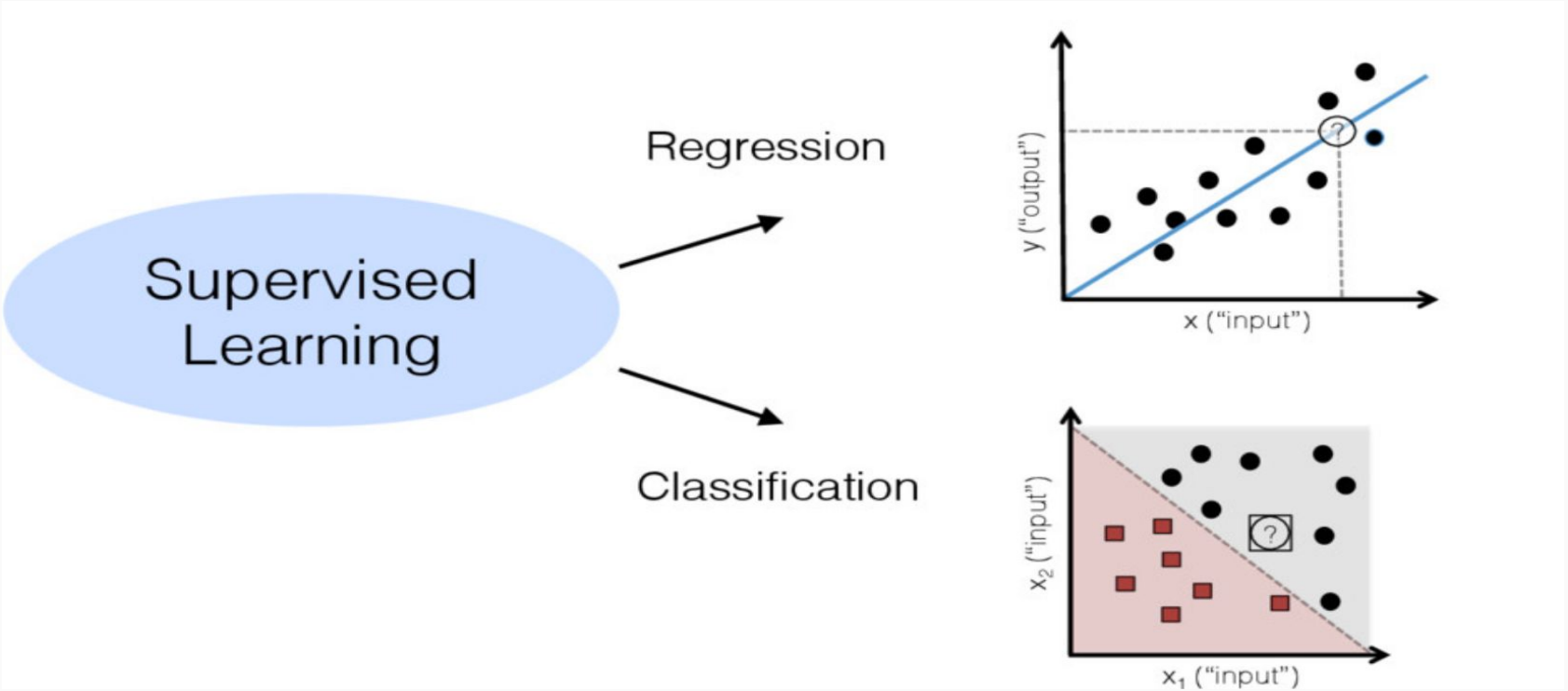
Week 5

Overview

Classification

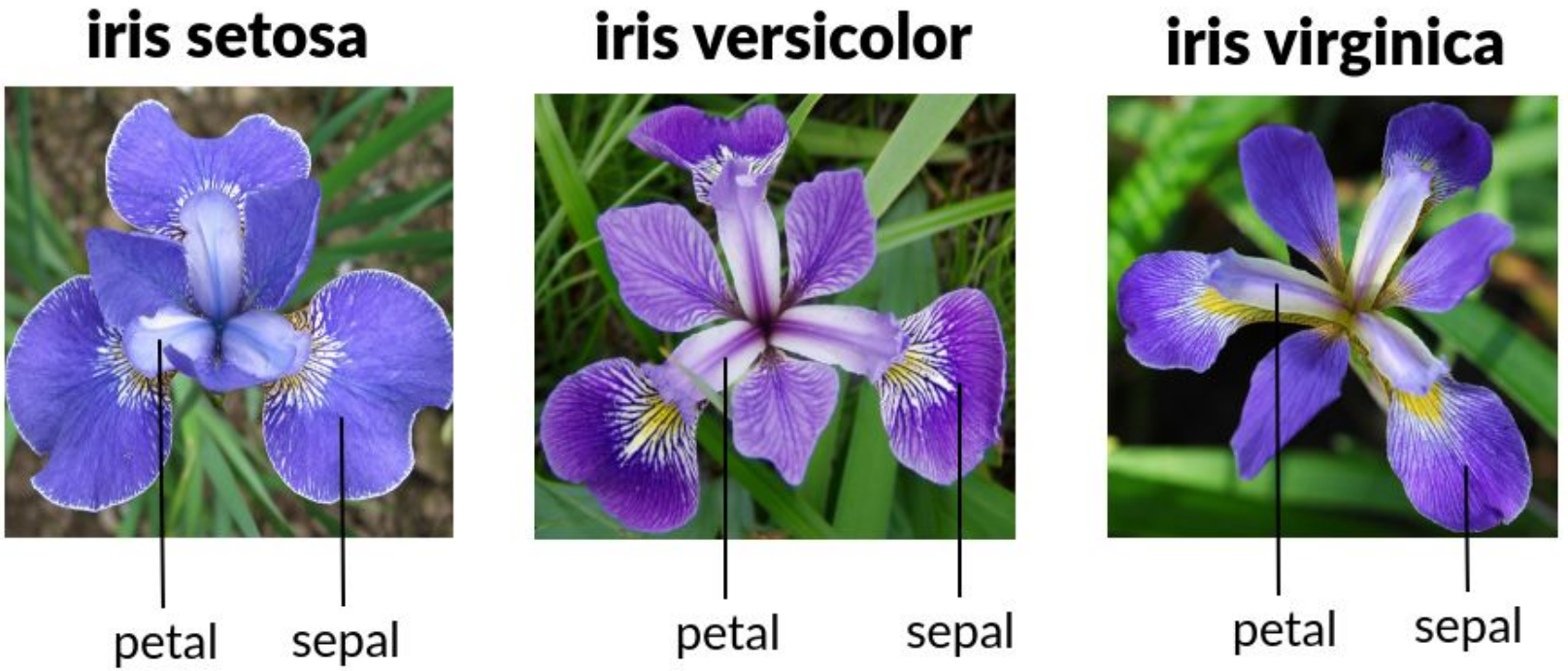
Hands-on

Classification



Credit:https://speakerdeck.com/rasbt/learning-scikit-learn-an-introduction-to-machine-learning-in-python-at-pydata-chicago-2016

Classification - Iris Dataset



Credit: https://thegoodpython.com/iris-dataset/

UCI Machine Learning Repository

Features in Iris Dataset

features (columns)

$\mathbf{X} =$

	sepal length [cm]	sepal width [cm]	petal length [cm]	petal width [cm]
1	5.1	3.5	1.4	0.2
2	4.9	3.0	1.4	0.2
50	6.4	3.5	4.5	1.2
	.			
	.			
	.			
150	5.9	3.0	5.0	1.8

samples (rows)

$\mathbf{y} =$

setosa
setosa
versicolor
.
.
.
virginica

Credit:https://speakerdeck.com/rasbt/learning-scikit-learn-an-introduction-to-machine-learning-in-python-at-pydata-chicago-2016

Few Common Classifiers

Perceptron

Naive Bayes

Decision Tree

K-Nearest Neighbor

Logistic Regression

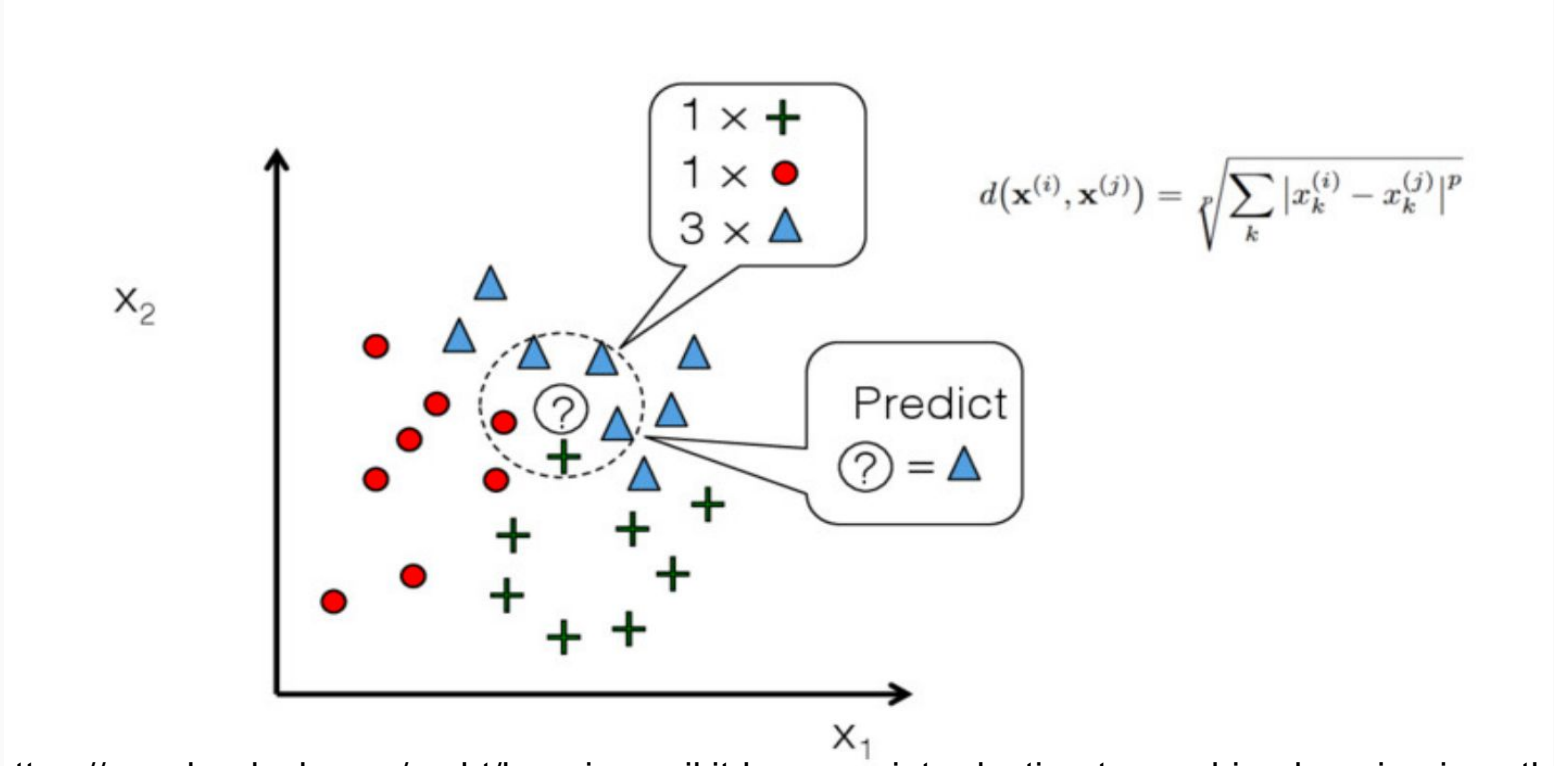
Artificial Neural Network / Deep Learning

Support Vector Machine

Ensemble Methods: Random Forest, Bagging, AdaBoost

Credit:https://speakerdeck.com/rasbt/learning-scikit-learn-an-introduction-to-machine-learning-in-python-at-pydata-chicago-2016

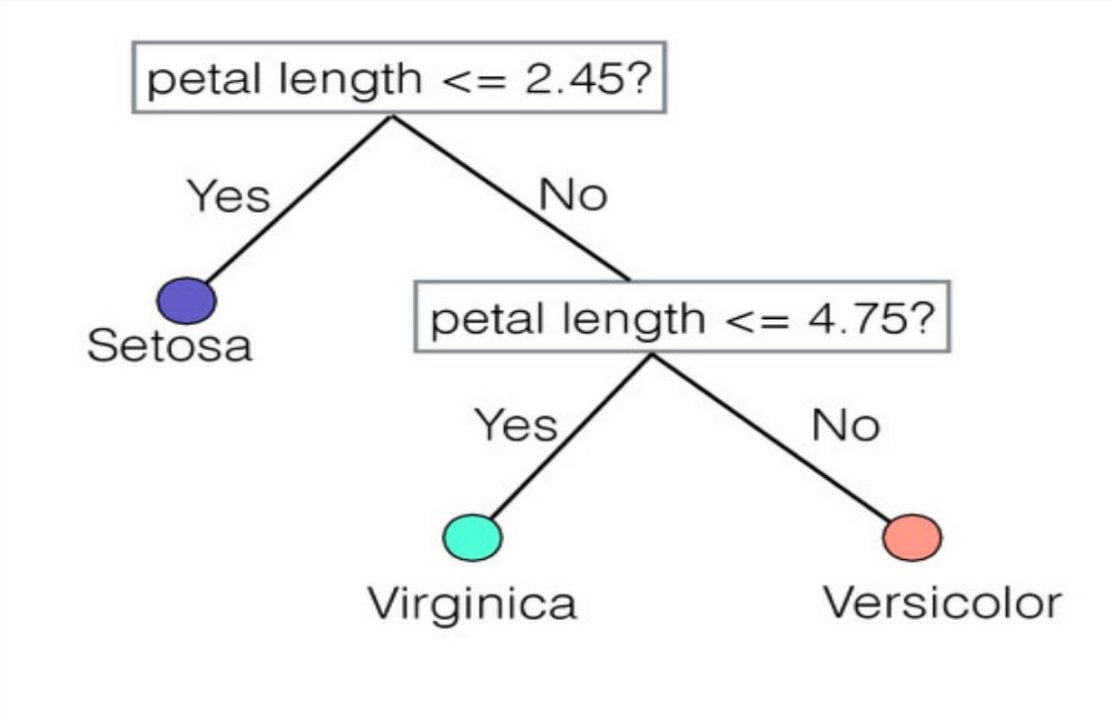
K-NN Nearest Neighbour Classification



$$d(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \sqrt[p]{\sum_k |x_k^{(i)} - x_k^{(j)}|^p}$$

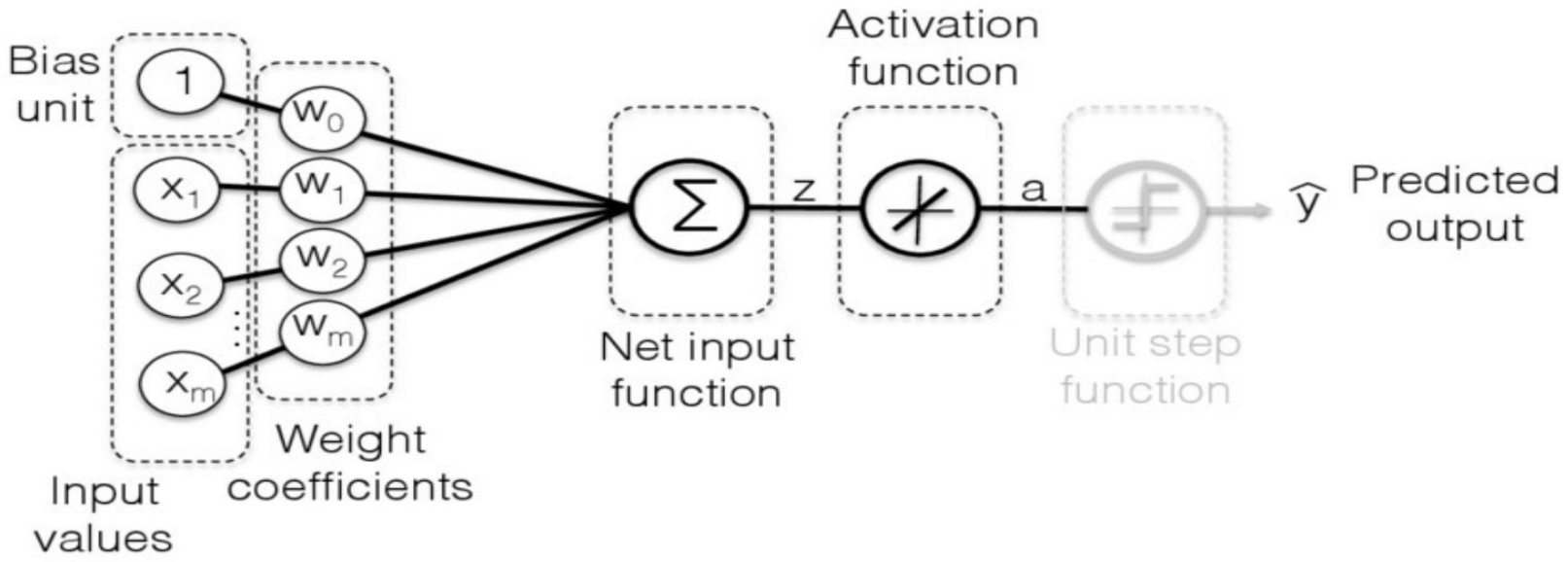
Credit:https://speakerdeck.com/rasbt/learning-scikit-learn-an-introduction-to-machine-learning-in-python-at-pydata-chicago-2016

Decision Tree



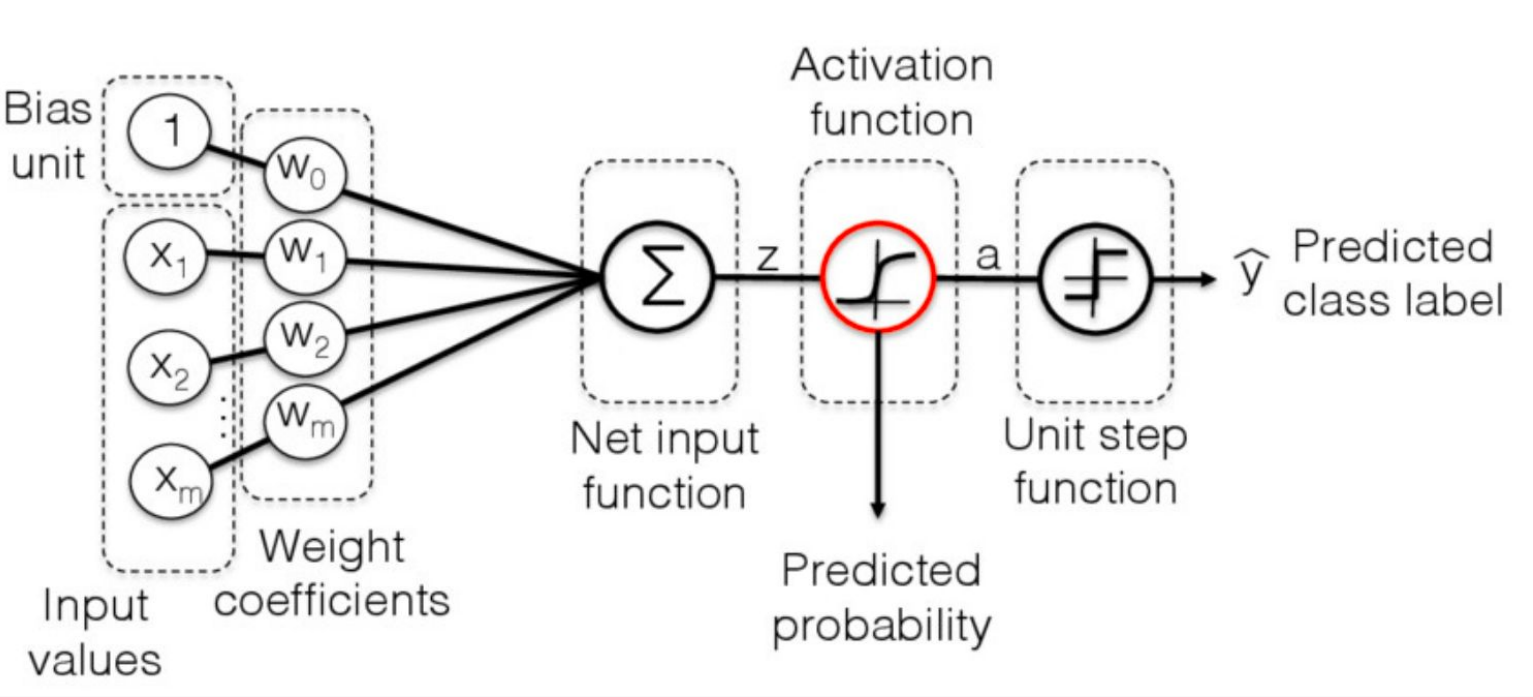
Credit:https://speakerdeck.com/rasbt/learning-scikit-learn-an-introduction-to-machine-learning-in-python-at-pydata-chicago-2016

Linear Regression -- recap



Credit:https://speakerdeck.com/rasbt/learning-scikit-learn-an-introduction-to-machine-learning-in-python-at-pydata-chicago-2016

Logistic Regression



Credit:https://speakerdeck.com/rasbt/learning-scikit-learn-an-introduction-to-machine-learning-in-python-at-pydata-chicago-2016

Scikit-learn API

```
class SupervisedEstimator(...):
    def __init__(self, hyperparam, ...):
        ...
    def fit(self, X, y):
        ...
        return self
    def predict(self, X):
        ...
        return y_pred
    def score(self, X, y):
        ...
        return score
    ...
```

Credit:https://speakerdeck.com/rasbt/learning-scikit-learn-an-introduction-to-machine-learning-in-python-at-pydata-chicago-2016

Categorical Variables

	color	size	prize	class
0	green	M	10.1	class1
1	red	L	13.5	class2
2	blue	XL	15.3	class1

nominal

green → (1,0,0)
red → (0,1,0)
blue → (0,0,1)

ordinal

M → 1
L → 2
XL → 3

	class	color=blue	color=green	color=red	prize	size
0	0	0	1	0	10.1	1
1	1	0	0	1	13.5	2
2	0	1	0	0	15.3	3

Credit:https://speakerdeck.com/rasbt/learning-scikit-learn-an-introduction-to-machine-learning-in-python-at-pydata-chicago-2016

Hands-On

Week 5 Jupyter NoteBook:

Next Week

Neural Networks and Clustering

References