



Machine Learning For Breast Cancer Detection

Behjat Saeidian
Python
Professor: Mr. Doozandeh



Breast Cancer And Machine Learning?

Outline



- What is breast cancer?
- What is machine learning?
- Wisconsin dataset
- Evaluate metrics
- K-Nearest Neighbors algorithm
- Random Forest algorithm
- Naive Bayes algorithm
- Support vector Machine algorithm
- Summary
- References



What is Breast Cancer?

Breast Cancer



- cells in the breast grow out of control
- The most frequently occurring cancer among women
- Early detection increase chances of survival
- Mammogram is a medical Technique for detection



What is Machine Learning?

Machine Learning



- Branch of Artificial Intelligence
- Inculcates the ability of learning into system
- Supervised learning
- Unsupervised learning



Wisconsin Dataset

Wisconsin Dataset



Dataset	No. of Attributes	No. of Instances	No. of Classes
Wisconsin Diagnosis Breast Cancer(WDBC)	32	569	2

- 357 benign instance
- 212 malignant instance



Evaluation Metrics

Evaluation Metrics



- Accuracy :
$$\frac{TP + TN}{TP + FP + TN + FN}$$

- Recall :
$$\frac{TP}{TP + FN}$$

- Precision :
$$\frac{TP}{TP + FP}$$

- F1 Score :
$$\frac{2 * (\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}}$$

TP : True positive
TN : True negative
FP : False Positive
FN : False Negative



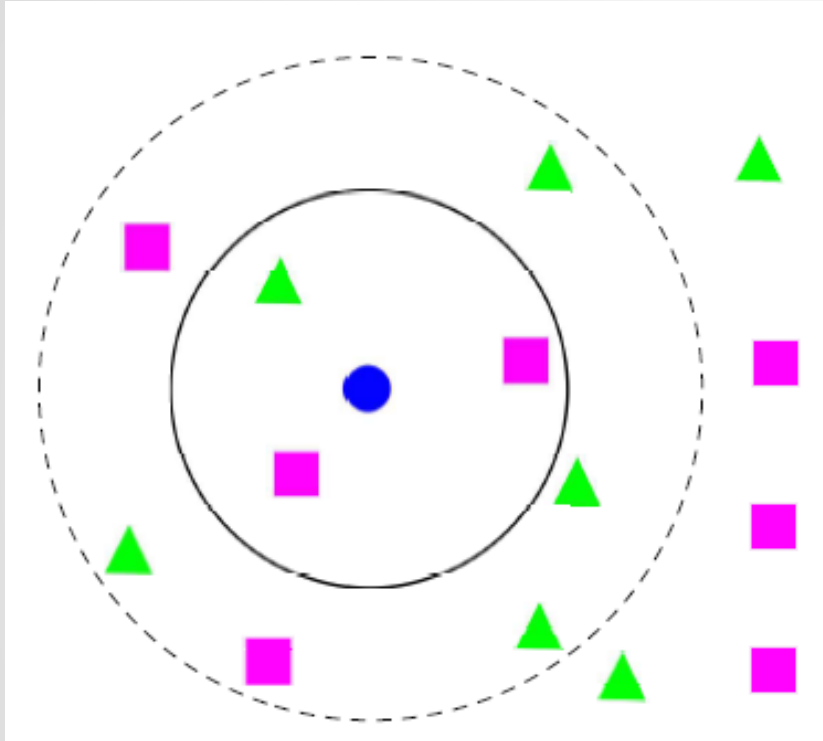
K-Nearest Neighbors (KNN)

What is KNN?



- Supervised machine learning algorithm
- Calculate distance to other data
 - Euclidean distance , Manhattan distance, Hamming distance, ...
- Find K-nearest neighbors
- Set label for data
 - Classification : mode of K labels
 - Regression : mean of K labels

KNN Classification



- Blue circle is test sample
- Green triangle means malignant
- Pink square means benign

- Different k , Different labels
- Run algorithm many times with different k to find optimal k



KNN Results

		Predicted	
		Benign	Malignant
Actual	Benign	107	1
	Malignant	6	57

- Accuracy = 95.9%
- Precision = 98.27%
- Recall = 90.47%
- F1 Score = 94.20%



Random Forest

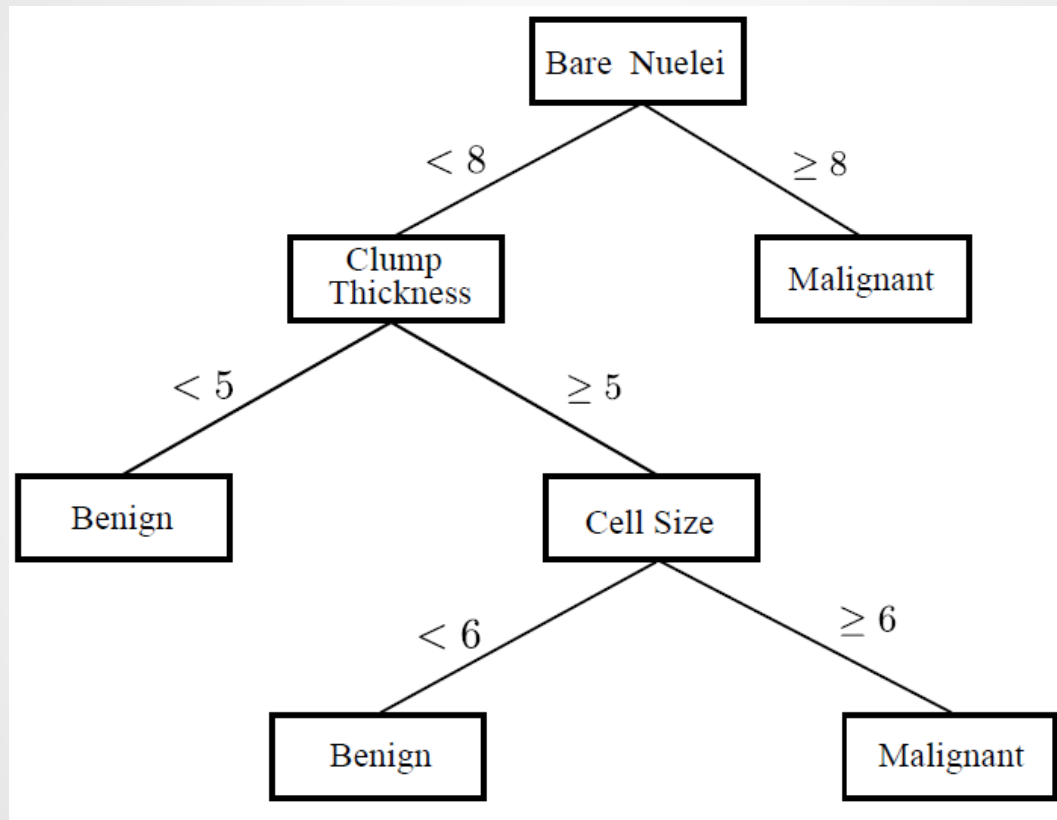
What is Random Forest?



- Supervised machine learning algorithm
- Large number of individual Decision Tree
- Majority – Voting or averaging on decision trees results
- Use for Classification and Regression

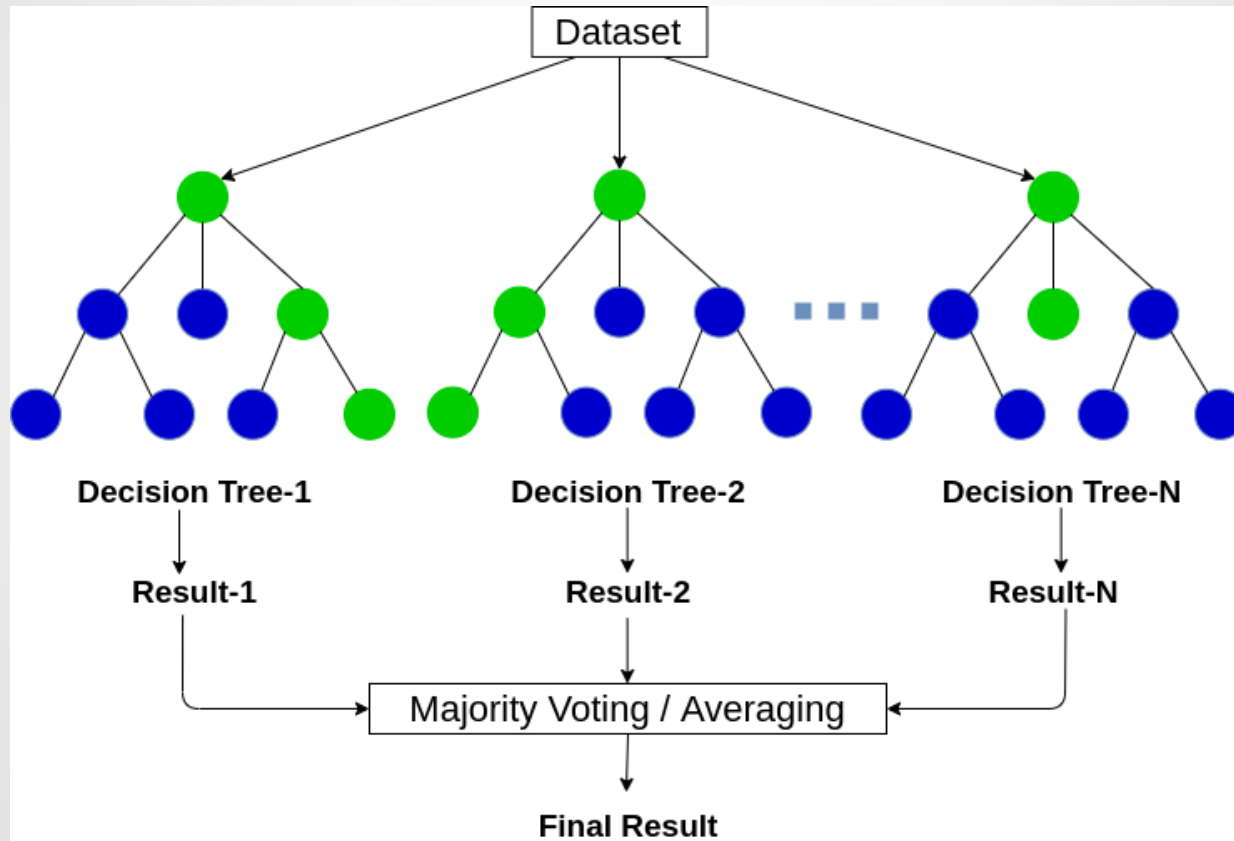


Decision Tree



Benign or malignan?

Random Forest



Random Forest Results



		Predicted	
		Benign	Malignant
Actual	Benign	103	5
	Malignant	4	59

- Accuracy = 94.74%
- Precision = 92.18%
- Recall = 93.65%
- F1 Score = 92.90%



Naive Bayes

What is Naive Bayes?



- Probabilistic supervised machine learning model
- Use for classification
- Based on Bayes theorem

$$P(y | X) = \frac{P(X | y) \cdot P(y)}{P(X)}, X = (x_1, x_2, x_3, x_4, \dots, x_n)$$

- Types of naive Bayes Classifier
Multinomial, Bernoulli, Gaussian



Naive Bayes Results

		Predicted	
		Benign	Malignant
Actual	Benign	101	7
	Malignant	9	54

- Accuracy = 94.47%
- Precision = 88.52%
- Recall = 85.71%
- F1 Score = 87.09%



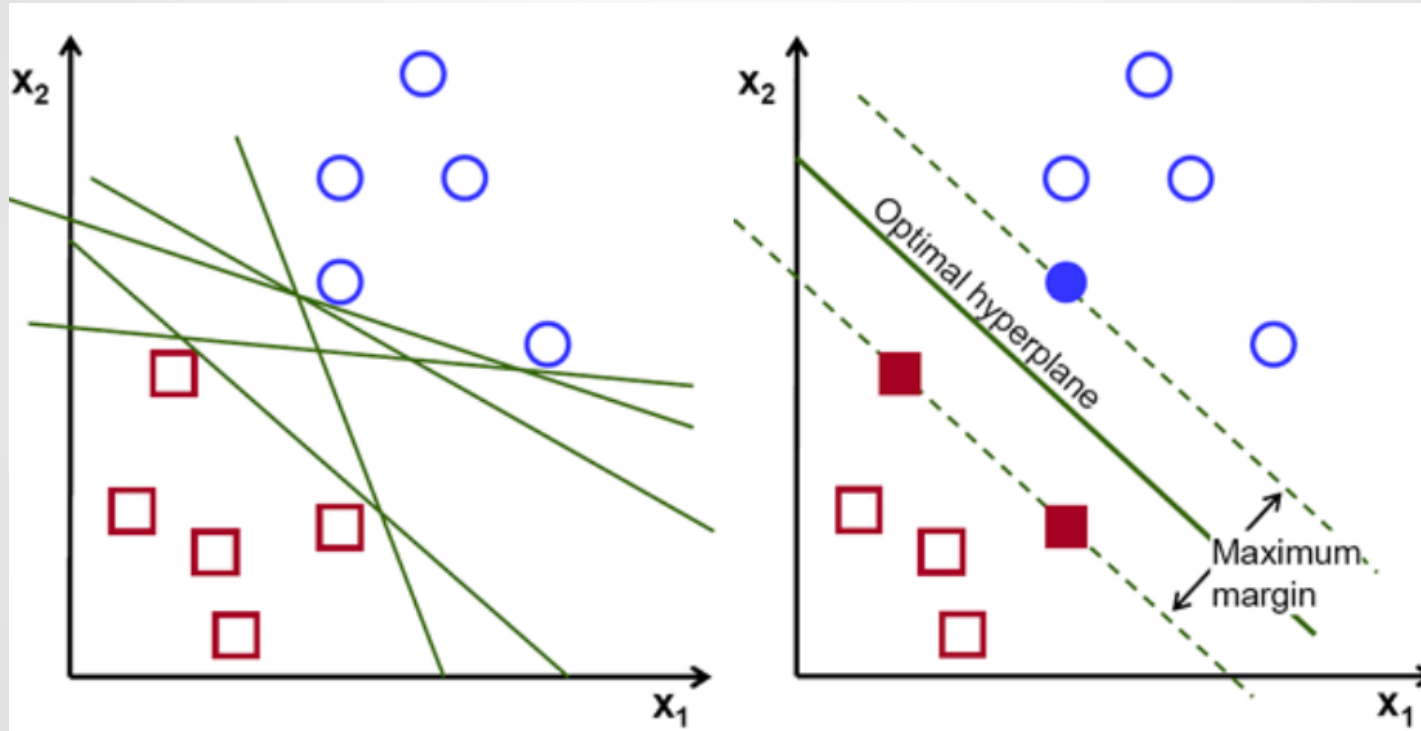
Support Vector Machine (SVM)

What is SVM?



- Supervised machine learning algorithm
- Use for Classification and regression
- Find hyperplane for separate classes
- Set label for data

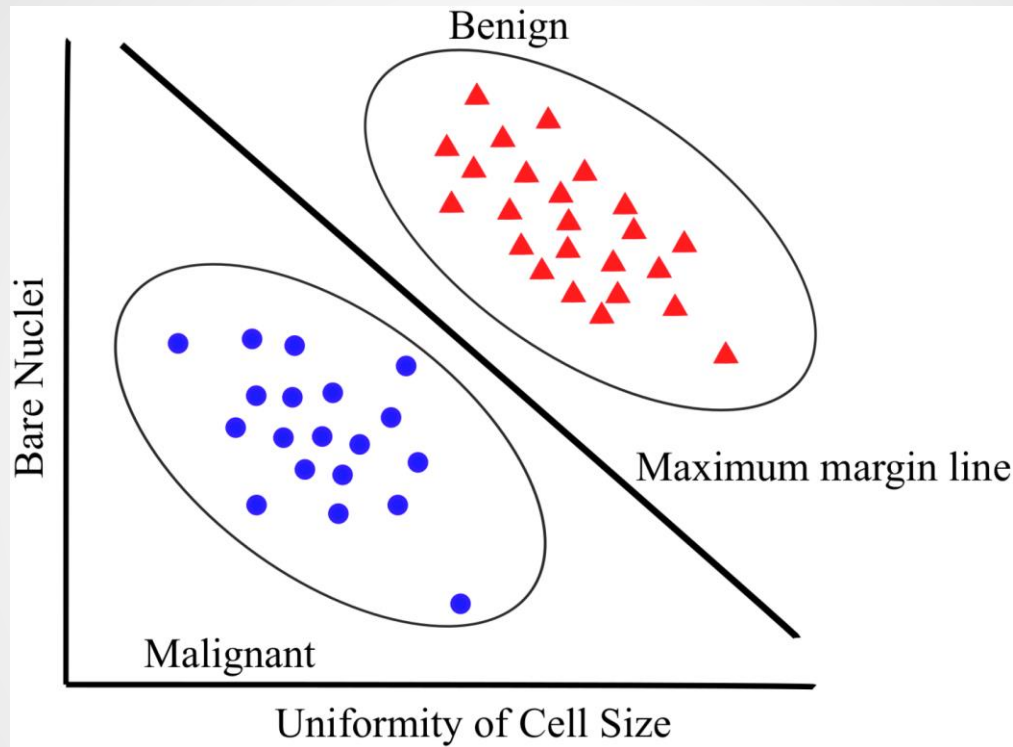
SVM



Optimal Hyperplane has maximum margin



SVM Results



- Accuracy = 97.0%
- Precision = 96.55%
- Recall = 96.85 %
- F1 Score = 96.69%

Summary



- Early detection of breast cancer can save a lot of lives
- Compare four supervised machine learning algorithm on Wisconsin dataset for Breast Cancer detection
- KNN, Random Forest, Naive Bayes on same condition
 - KNN has best accuracy(95.90) , precision(98.27) , F1 Score(94.2)
 - Random Forest(93.65) has best recall
- SVM has 97% accuracy , 96.55% precision , 96.85% recall and 96.69% F1 Score

References



- [1] Shubham Sharma, Archit Aggarwal, Tanupriya Choudhury, “Breast Cancer Detection Using Machine Learning Algorithms”,2018
- [2] Dana Bazazeh, Raed Shubairm, “Comparative Study of Machine Learning Algorithms for Breast Cancer Detection and Diagnosis”,2016
- [3] Wenbin Yue, Zidong Wang, Hongwei Chen, Annette Payne, Xiaohui Liu, “Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis”,2018



Thank You 😊

Questions?