

# Support Vector Machines

---



Prepared By: Rana Khan

## Rational Statement

Mr. John Hughes would like you to revisit the cancer.csv dataset and create a standard and **optimized SVM model** for desired outcomes.

# Independent Variable

- ID - ID number
- Clump Thickness - 1-10
- UofCSize - Uniformity of Cell Size 1-10
- UofShape - Uniformity of Cell Shape 1-10
- Marginal Adhesion - 1-10
- SECSize - Single Epithelial Cell Size 1-10
- Bare Nuclei - 1-10
- Bland Chromatin - 1-10
- Normal Nucleoli - 1-10
- Mitoses - 1-10

# Dependent Variable

- Benign (i.e. No Cancer) - 2
- Malignant (i.e. Cancer) - 4

# Navigation Synopsis



Copy cancer.csv  
into Pythondata2204 directory.  
Ensured the file is called  
cancer.csv



Launched Jupyter  
NoteBook



Navigated to  
Pythondata2204  
Directory



Created a new Python  
NoteBook by clicking on  
“New Drop Down and  
Choose “Python3”



Renamed  
NoteBook by  
Clicking on File->Rename



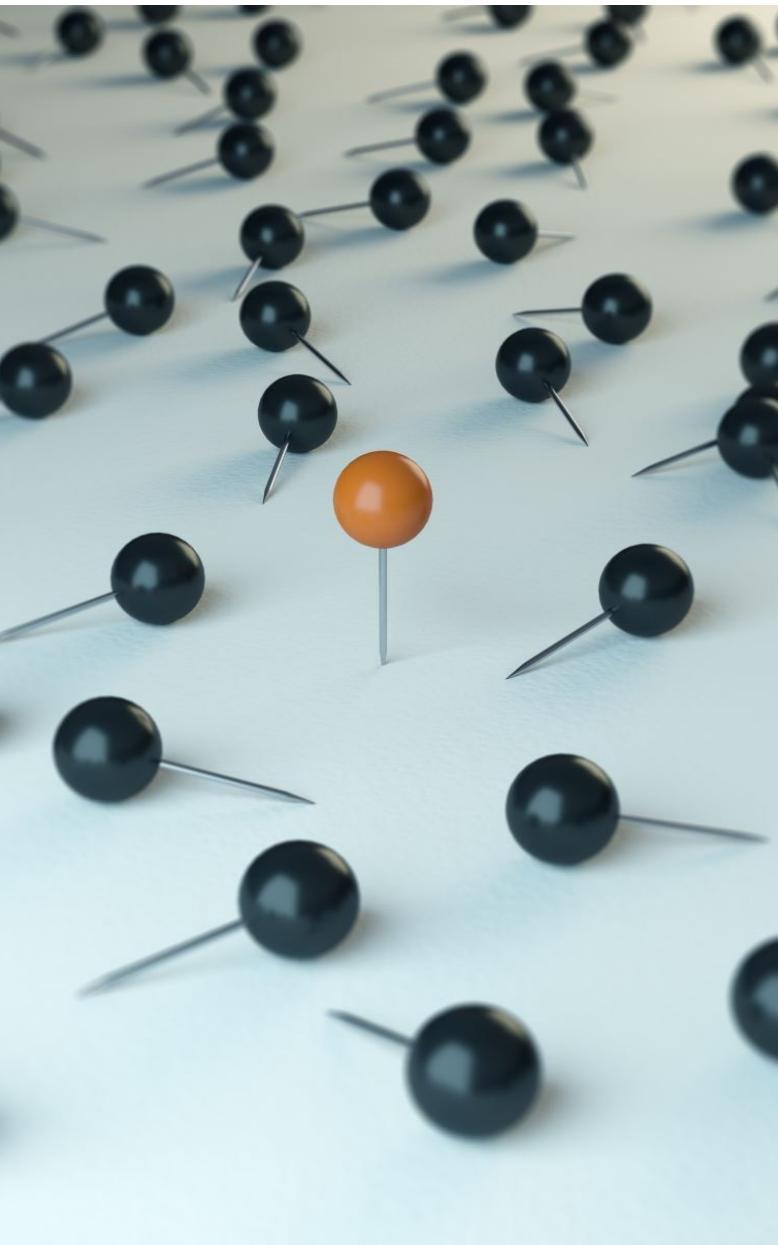
Entered “Assignment #5-  
Sipport Vector Machines”  
and  
Clicked OK

# Learning Curve for Original SVM Model

---



# Learning Curve for Original SVM Model and Key Insights



Model Evaluation – Accuracy Score  
Support Vector Machines **0.96** +/- 0.03

1. To establish learning curve, we went through linear, polynomial, rbf, and sigmoid also the gamma function is very low/pretty small number.
2. To get the most out of the model, we made hyperparameter tuning which the help of best kernel strength. We used *auto function from SVM* to attain the best outcomes.
3. Training and validation sets are close to each other. Model accuracy we attained is **96%** which is pretty great with the given range.

# Optimized SVM Model Classification Model Report



# Optimized SVM Model Classification Model Report

---

## Confusion Matrix

```
[ [86  3]
 [ 0 48]]
```

## Classification Report

		precision	recall	f1-score	support
	2	1.00	0.97	0.98	89
	4	0.94	1.00	0.97	48
	accuracy			0.98	137
	macro avg	0.97	0.98	0.98	137
	weighted avg	0.98	0.98	0.98	137

# Optimized SVM Model Classification Model Report and Key Insights

---



## Accuracy:

Accuracy is the ratio of correct predictions to total predictions made. In this case, accuracy is **98%**

## Precision:

Precision is about being precise, i.e., *how accurate your model is*. In other words, we can say, when a model makes a prediction, how often it is correct. In this case, precision is **97%**

## Recall:

The recall is the ratio of correctly predicted positive observations to all observations in the actual class. In this case, recall **98%**

## F1 Score:

With the predictions and validations, C parameter using 10, gamma parameter 0.1 with polynomial kernel. Our F1 score is **98 %** which is highly ideal.



# Recommendations

Based on the outcome of the model, the followings are the **recommendations** which can be furnished to Mr. John Hughes:

- Even though SVM model performed quite well with the accuracy of **98%** BUT to further enhance the model performance, more data can be requested
- Perform with another set of algorithms
- Adapt the methodology of transformation