# Artificial Intelligence / SVM Project / Documentation Yazdan Mastery Farahani / 99522059

#### 1. Linear Kernel

### • Usage:

- When the data is linearly separable.
- For large datasets, it is computationally less expensive compared to other kernels.
- When the number of features is large compared to the number of samples.

### 2. Polynomial Kernel

#### Usage:

- When the data has polynomial relationships.
- For moderately complex relationships where the linear kernel fails.
- Tuning the degree dallows control over the flexibility of the decision boundary.

### 3. Radial Basis Function (RBF) Kernel / Gaussian Kernel

#### Usage:

- When there is no prior knowledge about the data.
- Effective for non-linearly separable data.
- Versatile and can handle complex decision boundaries.
- Suitable when the relationship between features is not linear.

### 4. Sigmoid Kernel

#### Usage:

- Similar to a two-layer neural network.
- For certain types of datasets where sigmoid function characteristics are beneficial.
- Rarely used in practice compared to other kernels.

# Part-1. Spam Emails

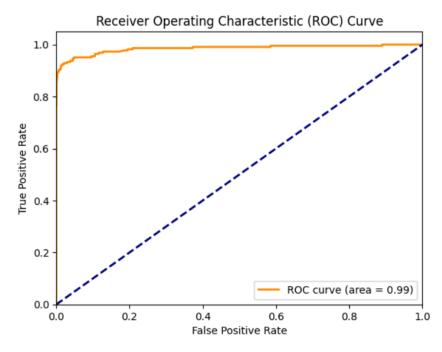
linear : 0.98 poly : 0.87 rbf : 0.87 sigmoid : 0.87

Best Kernel: Linear

Reason:

High Dimensionality of Text Data Simplicity Nature of the Dataset

Accuracy: 0.9 Classification		recall	f1-score	support
0	0.98	1.00	0.99	1448
1	0.98	0.87	0.92	224
accuracy			0.98	1672
macro avg	0.98	0.93	0.96	1672
weighted avg	0.98	0.98	0.98	1672
Confusion Mat [[1445 3] [ 29 195]]				



# Part-2. Diabetes

linear : 77 % rbf : 79 % sigmoid : 53 %

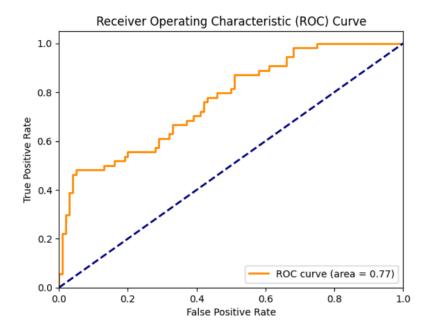
Best Kernel: RBF

Reason:

Non-linear Relationships Automatic Adaptation

\*Poly Kernel (Not Working): Likely due to high computational complexity and potential overfitting, leading to very slow or impractical training times.

Accuracy: 0.7 Classificatio		recall	f1-score	support
	pr cc151011		11 30010	опррот с
0	0.77	0.95	0.85	100
1	0.84	0.48	0.61	54
accuracy			0.79	154
macro avg	0.81	0.72	<b>0.7</b> 3	154
weighted avg	0.80	0.79	0.77	154
Confusion Mat [[95 5] [28 26]]	rix:			



# **Part-3. Boston Housing**

Selected Kernel: RBF

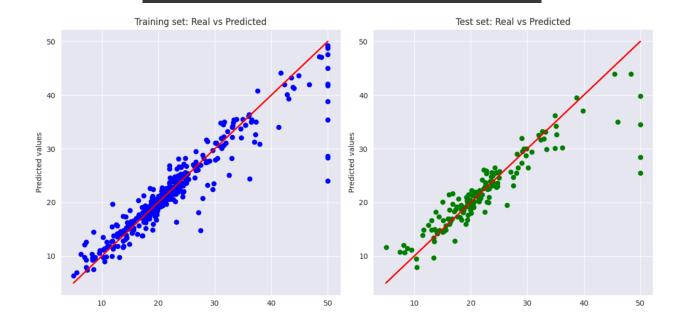
Reason:

Non-linear Relationships Automatic Adaptation

CV: 0.782600693247584

R2\_score (train): 0.8742715669147498 R2\_score (test): 0.8044070157723574

3.8388977809328355



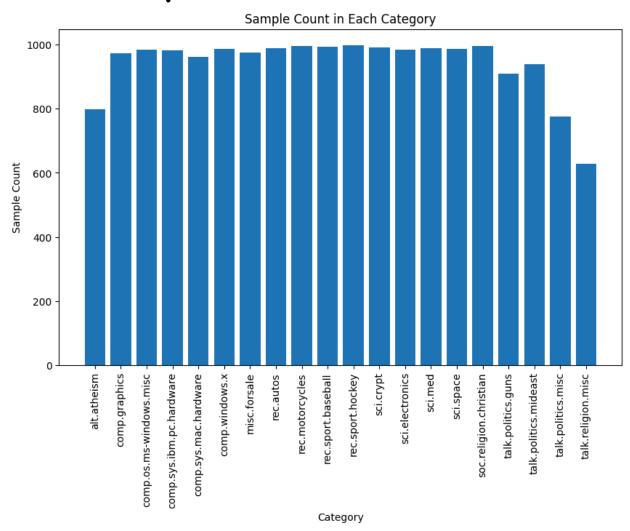
## Part-4. 20NewsGroup

Best Kernel: Linear

Reason:

SimplicityHigh Dimensionality of Text DataNature of the Dataset

linear : 69% sigmoid : 67% poly: 11%



SVM Accuracy: 91.751% SVM Recall: 91.508%

SVM F1: 91.650%

C	conf	usio	n M	atrix
0	0	1	0	0

									C	OH	usic	111 141	auı	X									
	alt.atheism -	157	2	0	1	0	0	0	0	0	1	0	0	1	2	0	5	0	0	0	6		- 200
	comp.graphics -	0	171	4	6	1	10	4	0	0	1	0	0	1	1	0	0	0	0	0	0		200
	comp.os.ms-windows.misc -	0	9	187	16	3	3	1	1	0	0	0	0	0	0	0	0	0	0	1	0		
	comp.sys.ibm.pc.hardware -	0	4	4	153	2	2	4	0	0	0	0	0	10	0	0	0	0	0	0	0		- 175
	comp.sys.mac.hardware -	0	2	1	8	156	1	5	0	0	0	0	0	3	0	1	0	0	0	0	0		
	comp.windows.x -	0	6	5	0	0	166	1	0	0	0	0	0	1	0	0	0	0	0	0	0		- 150
	misc.forsale -	0	1	2	3	3	0	184	4	0	1	0	1	4	1	0	1	0	0	0	0		
	rec.autos -	0	1	1	1	0	1	7	210	2	0	0	0	5	0	0	0	0	0	0	0		
_	rec.motorcycles -	0	0	0	0	1	0	4	3	172	2	0	0	1	0	0	0	0	0	0	0		- 125
True Label	rec.sport.baseball -	0	2	0	2	0	0	0	0	0	187	4	0	1	0	0	1	0	0	0	0		
rue	rec.sport.hockey -	0	1	0	0	0	0	2	0	0	2	199	0	0	0	0	0	0	0	0	0		- 100
_	sci.crypt -	0	3	0	0	0	3	0	0	0	1	0	208	1	0	0	0	0	0	2	0		
	sci.electronics -	0	2	0	2	2	1	3	0	0	0	0	0	160	2	0	0	0	0	0	0		- 75
	sci.med -	0	1	0	1	1	1	3	0	1	0	0	0	3	185	2	0	1	0	1	0		,,,
	sci.space -	0	6	0	0	0	0	1	0	0	0	0	0	0	2	188	0	0	0	1	0		
	soc.religion.christian –	1	2	0	2	0	0	0	0	0	1	0	0	2	0	0	183	0	0	0	0		- 50
	talk.politics.guns -	1	1	2	0	0	2	2	0	0	0	0	0	0	1	0	0	163	0	1	0		
	talk.politics.mideast -	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	182	1	0		- 25
	talk.politics.misc -	1	3	0	0	0	0	1	1	0	0	0	0	1	2	0	1	3	1	157	1		
	talk.religion.misc -	3	0	0	0	0	1	0	1	0	0	0	0	1	1	0	8	6	1	2	91		- 0
		alt.atheism -	comp.graphics -	comp.os.ms-windows.misc -	comp.sys.ibm.pc.hardware -	comp.sys.mac.hardware -	comp.windows.x -	misc.forsale -	rec.autos -	rec.motorcycles -	rec.sport.baseball -	rec.sport.hockey -	sci.crypt -	sci.electronics -	sci.med -	sci.space -	soc.religion.christian -	talk.politics.guns –	talk.politics.mideast -	talk.politics.misc -	talk.religion.misc -		- 0
										PI E	CHICLE	-0112	Del										

Predicted Label

# Part5\_HealthExaminationSurvay

Accuracy Score: 0.963294538943599  Prediction: [0. 0. 0 0. 0. 0.]  Classification Report:											
	precision	recall	f1-score	support							
0.0	0.96	1.00	0.98	1076							
1.0	0.00	0.00	0.00	41							
accuracy			0.96	1117							
macro avg	0.48	0.50	0.49	1117							
weighted avg	0.93	0.96	0.95	1117							

