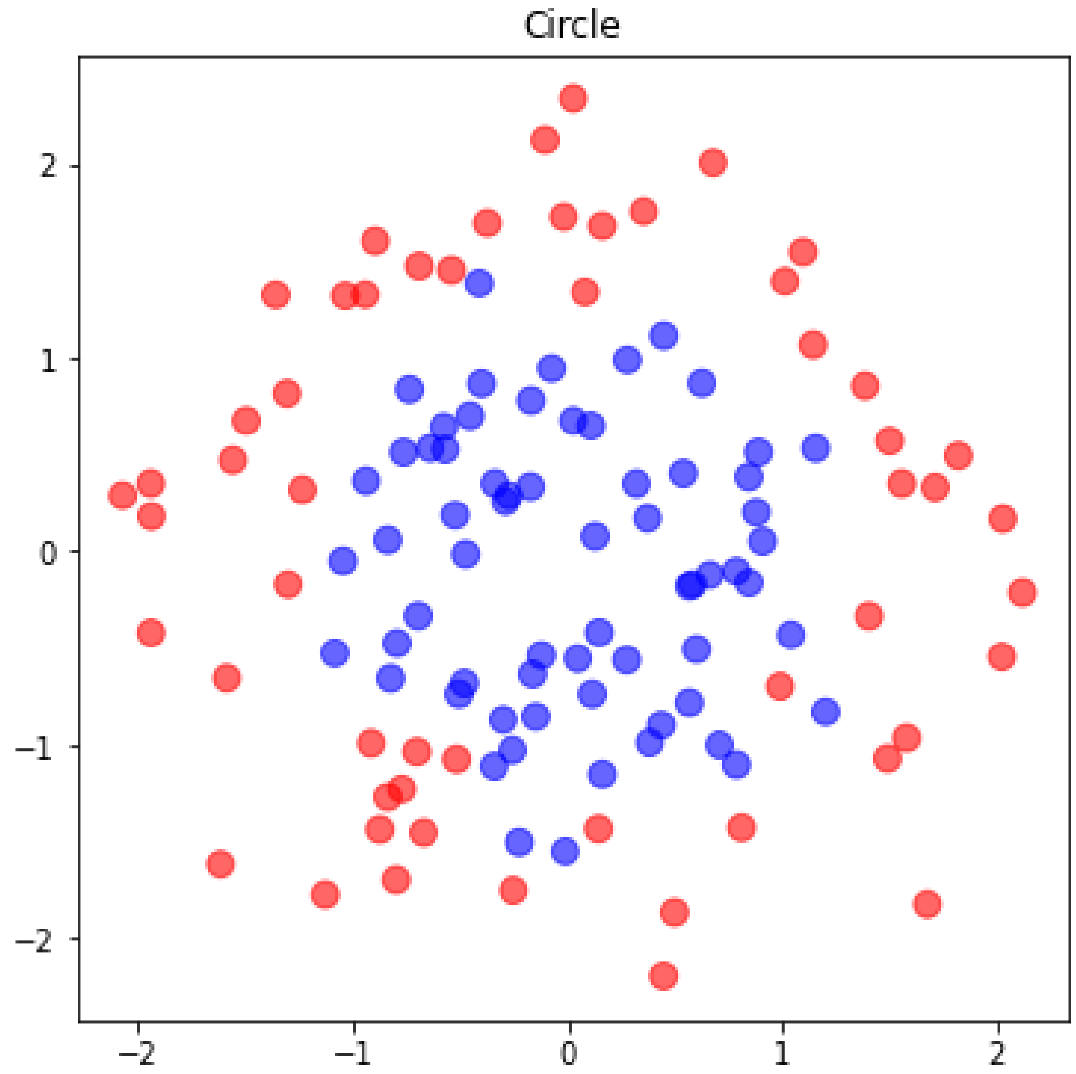
Support Vector Machine

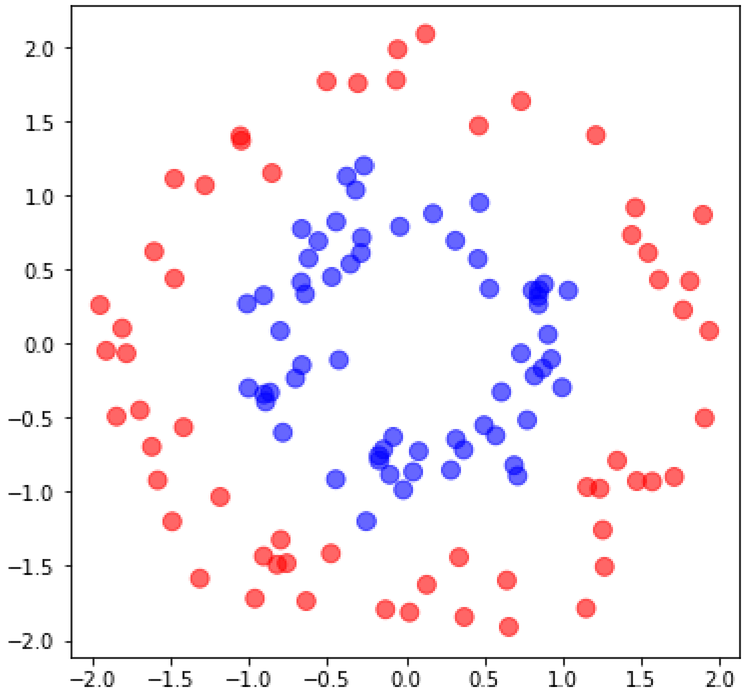
Comparison

What about this data?

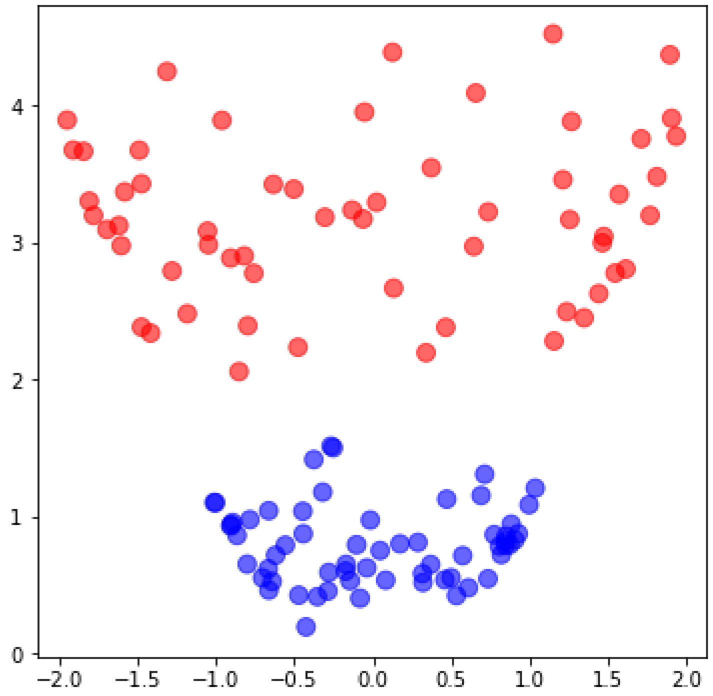


When data is not linearly separable

Not linearly separable in 2D



We can separate in 3D

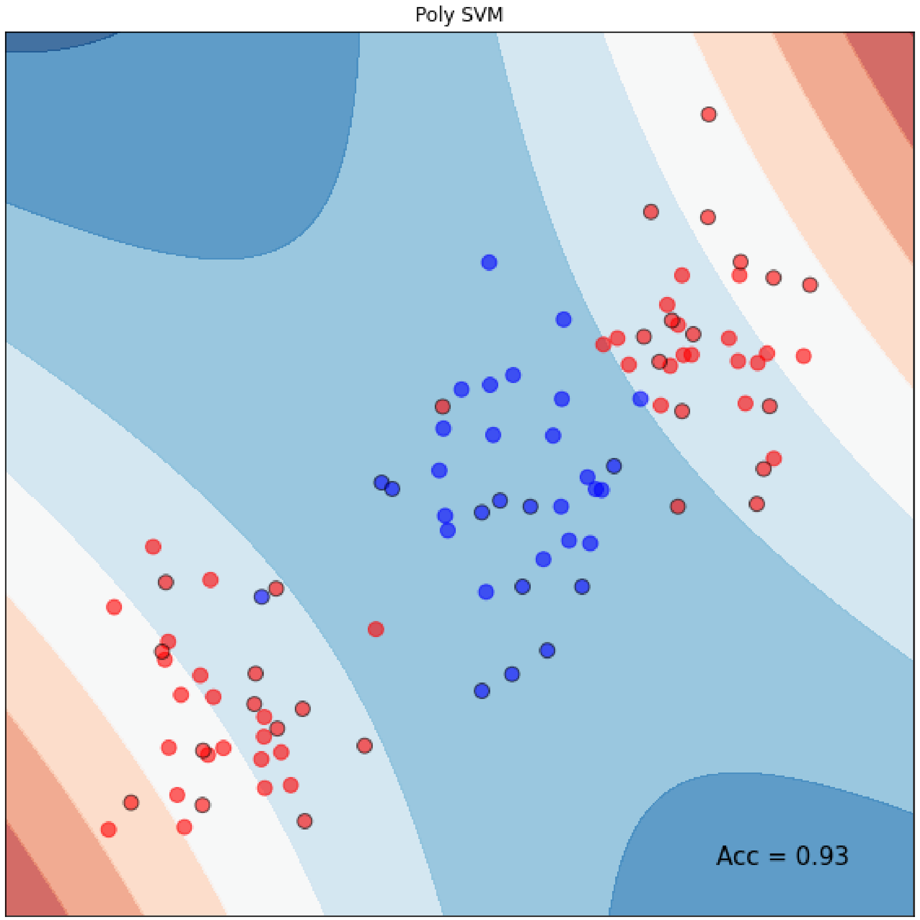


Polynomial Kernel

Non-linear kernels can take care of non-linear decision boundary

Polynomial kernel

$$K(x_i, x_{i'}) = (1 + \sum_{j=1}^{p} x_{ij} x_{i'j})^d$$

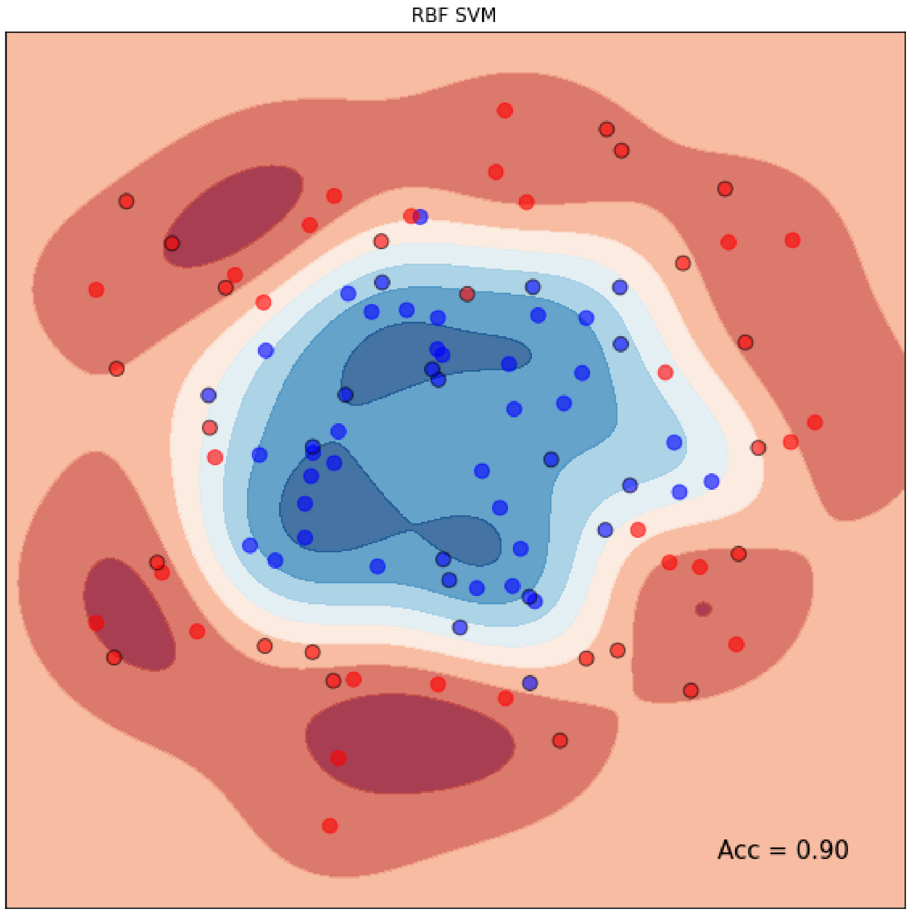


Radial Kernel

Non-linear kernels can take care of non-linear decision boundary

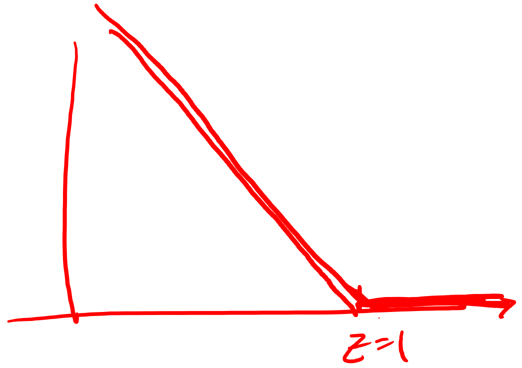
Radial Basis Function Kernel

$$K(x_i, x_{i'}) = \exp(-\gamma \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2)$$

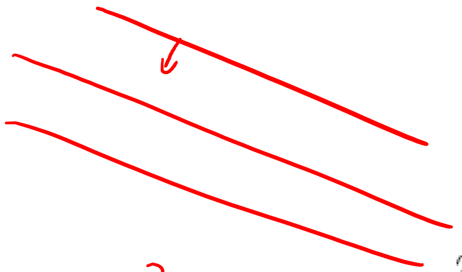


SVM's property

- Needs feature scaling
- Time complexity scales linearly to features
- Good for complex but small to medium dataset
- Loss function



C



$$\underset{\beta_0,\beta_1,\ldots,\beta_p}{\text{minimize}}\left\{\underbrace{\sum_{i=1}^{n}\max\left[0,1-\overset{z}{y_i f(x_i)}\right]}_{\text{Hinge loss}}+\lambda\underbrace{\sum_{j=1}^{p}\beta_j^2}\right\}$$

Usage

sklearn.svm.LinearSVC

liblinear

kernel X

C →

class sklearn.svm.LinearSVC(penalty='l2', loss='squared_hinge', *, dual=True, tol=0.0001, C=1.0, multi_class='ovr', fit_intercept=True, intercept_scaling=1, class_weight=None, verbose=0, random_state=None, max_iter=1000) [source]

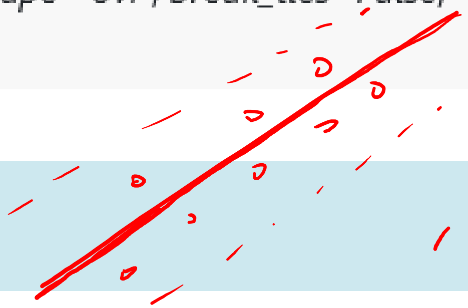
sklearn.svm.SVC

libsvm

kernel

poly

class sklearn.svm.SVC(*, C=1.0, kernel='rbf', degree=3, gamma='scale', coef0=0.0, shrinking=True, probability=False, tol=0.001, cache_size=200, class_weight=None, verbose=False, max_iter=- 1, decision_function_shape='ovr', break_ties=False, random_state=None) [source]



sklearn.svm.SVR

class sklearn.svm.SVR(*, kernel='rbf', degree=3, gamma='scale', coef0=0.0, tol=0.001, C=1.0, epsilon=0.1, shrinking=True, cache_size=200, verbose=False, max_iter=- 1) [source]

Usage

sklearn.svm.SVC

class `sklearn.svm.SVC`(**, C=1.0, kernel='rbf', degree=3, gamma='scale', coef0=0.0, shrinking=True, probability=False, tol=0.001, cache_size=200, class_weight=None, verbose=False, max_iter=- 1, decision_function_shape='ovr', break_ties=False, random_state=None*) [source]

SVM performance

Data 1.

- 13 features,
- 5200+ samples (4200 training)
- DT performance 0.61
- Sparse, and mostly categorical features

	sex	age	juv_fel_count	juv_misd_count	juv_other_count	priors_count	age_cat_25-45	age_cat_Greaterthan45	age_cat_Lessthan25	race_African-American
0	1	34.0	0.0	0.0	0.0	0.0	1	0	0	1
1	1	24.0	0.0	0.0	1.0	4.0	0	0	1	1
2	1	41.0	0.0	0.0	0.0	14.0	1	0	0	0
3	0	39.0	0.0	0.0	0.0	0.0	1	0	0	0
4	1	27.0	0.0	0.0	0.0	0.0	1	0	0	0

SVM performance

Data 2.

- 20 features,
- 5100+ samples (4100 training)
- DT performance 0.90
- Dense, real-valued features

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20
0	234.0	120.0	254.0	3260.0	70.0	1604.0	3397.0	8.0	2377.0	0.0	223.0	171.0	229.0	4406.0	3594.0	0.0	225.0	316.0	60.0	129.0
1	228.0	-9.0	251.0	876.0	164.0	3971.0	2991.0	7.0	2042.0	95.0	199.0	223.0	227.0	2107.0	3980.0	2.0	242.0	66.0	76.0	149.0
2	229.0	0.0	226.0	2765.0	39.0	3583.0	2766.0	10.0	842.0	0.0	201.0	187.0	221.0	4478.0	2297.0	0.0	218.0	113.0	175.0	131.0
3	248.0	45.0	221.0	4865.0	16.0	1031.0	2390.0	23.0	1104.0	371.0	254.0	225.0	193.0	937.0	1187.0	234.0	186.0	23.0	99.0	121.0
4	229.0	16.0	226.0	918.0	299.0	3488.0	2131.0	6.0	911.0	30.0	232.0	226.0	203.0	4568.0	1465.0	1.0	238.0	21.0	234.0	174.0

SVM performance

Data 3.

- 144 features,
- ~3000 samples (2400 training)
- DT performance 0.73
- Sparse, categorical features

	0	1	2	3	4	5	6	7	8	9	...	134	135	136	137	138	139	140	141	142	143
0	1	0	0	0	0	1	1	0	1	0	...	1	0	1	0	1	0	1	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	1	0	1	0	0	0	0
2	0	0	0	0	0	1	0	0	0	0	...	0	0	1	0	0	1	0	0	0	1
3	0	0	0	0	0	1	0	0	1	0	...	0	0	0	0	0	0	0	1	0	0
4	0	0	0	0	0	0	1	0	1	0	...	0	0	0	1	0	0	0	1	0	0

SVM performance

Data 4.

- 308 features,
- 5800+ samples (4700 training)
- DT performance 0.69
- Dense, real-valued features

	0	1	2	3	4	5	6	7	8	9	...	298	299	300	301	302	303
0	1.43676	0.000353	1972.84	0.007600	-0.015531	0.780376	103461	0.024912	-0.2213	47	...	36.5197	0.865036	0.359862	0.003133	2.76052	750.11
1	2.70690	0.000678	3955.21	-0.009876	-0.003330	0.733401	150978	0.028482	0.2259	55	...	64.3038	0.703948	0.610000	-0.006252	4.18539	1200.69
2	0.46472	0.000715	1453.81	0.002587	-0.005504	0.771300	88342	0.031772	0.0624	0	...	24.0798	0.806948	0.112245	-0.010088	1.31197	3876.00
3	3.06980	0.000357	2036.55	0.005065	0.009438	0.873578	123739	0.009533	-0.2277	38	...	44.5886	0.874592	0.348765	0.007156	2.19352	1013.80
4	1.65000	0.000545	2325.00	0.006807	-0.003393	0.803415	103812	0.014284	0.0149	14	...	30.0132	0.742672	0.284024	-0.003029	2.93776	1336.37

SVM performance

Data 5.

- 1776 features,
- 3750 samples (375 training)
- DT performance 0.70
- Some real-valued, mostly categorical (93%)

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	...	D1767	D1768	D1769	D1770	D1771	D1772
0	0.000000	0.497009	0.10	0.0	0.132956	0.678031	0.273166	0.585445	0.743663	0.243144	...	0.0	0.0	0.0	0.0	0.0	0.0
1	0.366667	0.606291	0.05	0.0	0.111209	0.803455	0.106105	0.411754	0.836582	0.106480	...	1.0	1.0	1.0	1.0	0.0	1.0
2	0.033300	0.480124	0.00	0.0	0.209791	0.610350	0.356453	0.517720	0.679051	0.352308	...	0.0	0.0	0.0	0.0	0.0	0.0
3	0.000000	0.538825	0.00	0.5	0.196344	0.724230	0.235606	0.288764	0.805110	0.208989	...	0.0	0.0	0.0	0.0	0.0	0.0
4	0.100000	0.517794	0.00	0.0	0.494734	0.781422	0.154361	0.303809	0.812646	0.125177		0.0	0.0	0.0	0.0	0.0	0.0

SVM performance - Accuracy

Model	Data 1 13/4200	Data 2 20/4100	Data 3 144/2400	Data 4 308/4700	Data 5 1776/375
DT (no reg)	0.613	0.898	0.727	0.690	0.675
RF	0.619 (n=500)	0.933 (n=1000)	0.809 (n=100)	0.760 (n=1000)	0.72 (0.80) (n=500)
GBM	0.666 (n=20)	0.943 (n=1000)	0.809 (n=500)	0.754 (n=200)	0.73 (0.80) (n=1000)
SVM	0.656 (C=0.2)	0.927 (C=1)	0.77-0.81 (C=0.5)	0.71-0.74 (C=2)	0.70 (0.79) (C=5)
Logistic (l2)	0.677	0.915	0.752*	0.706*	0.694*

SVM performance- Training Time

Model	Data 1 13/4200	Data 2 20/4100	Data 3 144/2400	Data 4 308/4700	Data 5 1776/375
DT (no reg)	26 ms	64 ms	42 ms	1.7 s	166 ms
RF	1.6 s (n=500)	8.6 s (n=1000)	331 ms (n=100)	63 s (n=1000)	1.2 s (n=500)
GBM	76 ms (n=20)	13.5 s (n=1000)	3.46 s (n=500)	81 s (n=200)	23 s (n=1000)
SVM	991 ms	532 ms	459 ms	5.05 s	1.37 s
Logistic (l2)	77 ms	37 ms	130 ms	334 ms	190 ms