Práctica 6: Support Vector Machines

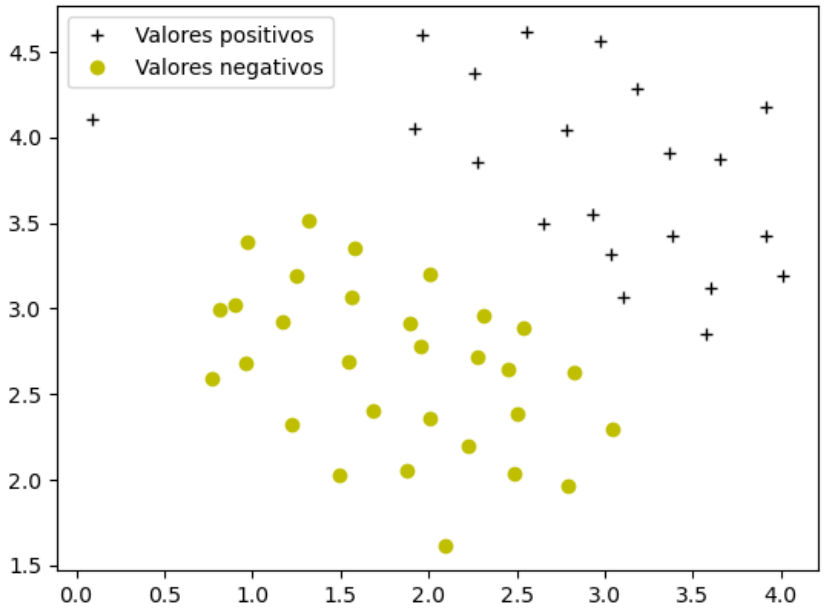
Nuria Bango Iglesias (nubango@ucm.es)

Álvar Julián de Diego López (alvarded@ucm.es)

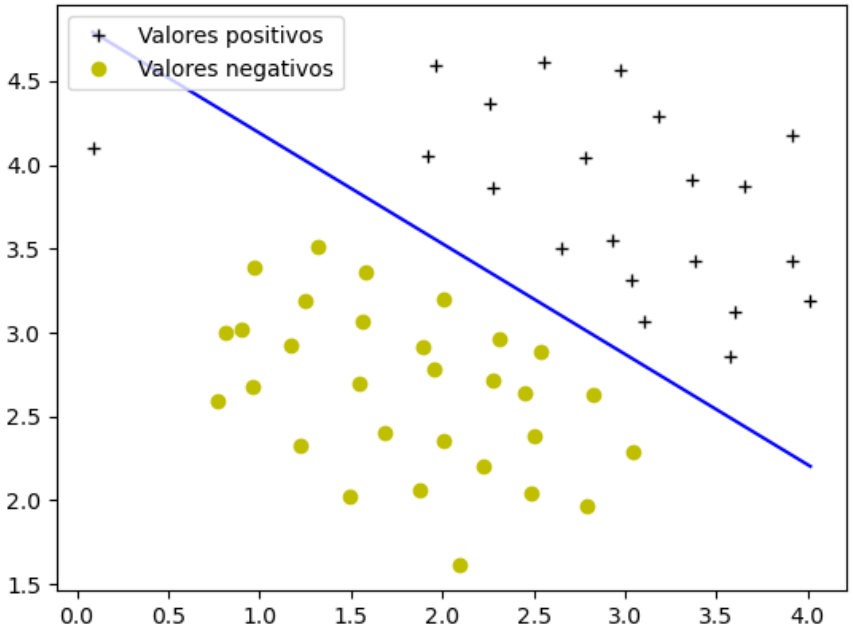
1.Support Vector Machines

1.1.Kernel lineal

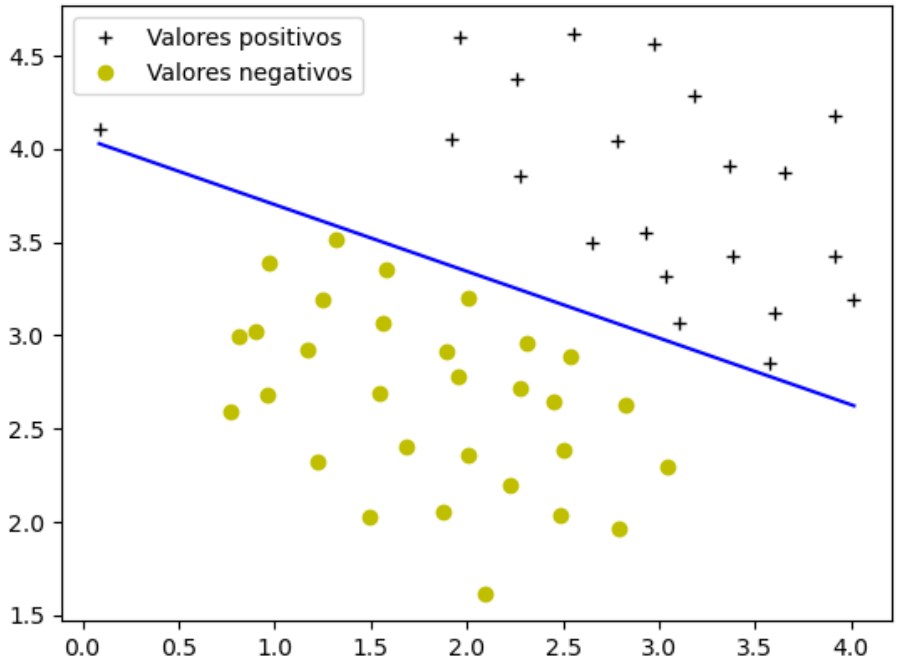
```
data1 = loadmat("ex6data1.mat")
X = data1['X']
y = data1['y']
Y_ravel = y.ravel()

print(X.shape)
plot_data(X, Y_ravel)

C = 1
model = svmLinearTrain(X, Y_ravel, C, 1e-3, -1)
visualizeBoundryLinear(X, Y_ravel, model)
C = 100
model = svmLinearTrain(X, Y_ravel, C, 1e-3, -1)
visualizeBoundryLinear(X, Y_ravel, model)
```



SVM con C = 1



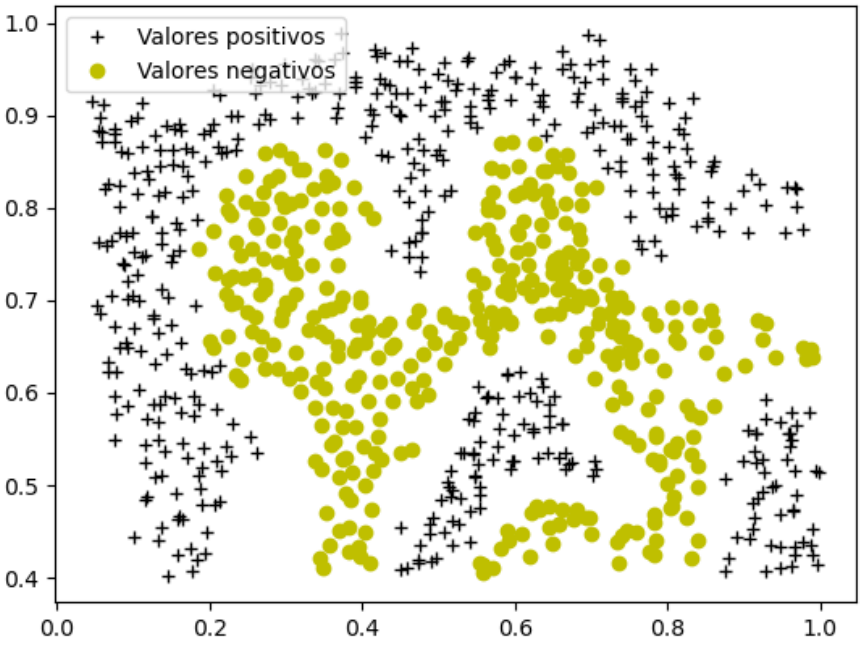
SVM con C = 100

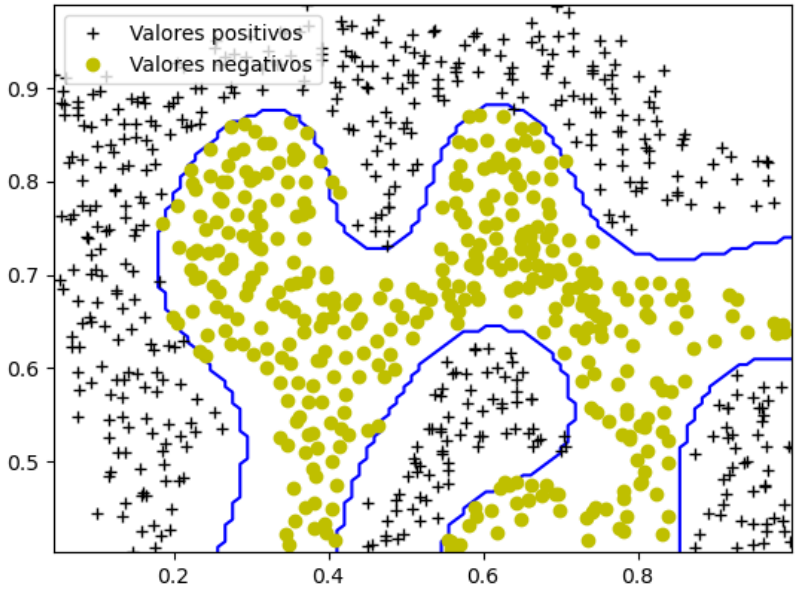


1.2.Kernel gaussiano

```
data2 = loadmat("ex6data2.mat")
X = data2['X']
y = data2['y']
Y_ravel = y.ravel()

plot_data(X, Y_ravel)
```

```
C = 1
sigma = 0.1
model = svmGaussianTrain(X, Y_ravel, C, 1e-3, 100, sigma)
visualizeBoundry(X, Y_ravel, model, sigma)
```





2.Detección de spam

```
email_contents = codecs.open('spam/0001.txt', 'r', encoding='utf-8', errors='ignore').read()
email_contents = preProcess(email_contents)
email_token_list = email2TokenList(email_contents)

vocab_dict = getVocabDict() # 1899 elementos
email_word_indices = tokenList2wordIndices(email_token_list, vocab_dict)
print(email_word_indices)

email_attributes = np.zeros((1899,), dtype=int)

for idx in email_word_indices:
    email_attributes[idx] = 1
```

```
(base) C:\Users\alvar\Documents\GitHub\AprendizajeAutomatico\Prácticas\Práctica 6>python practica6.py
[688, 531, 1689, 133, 1120, 1120, 1120, 1120, 1224, 425, 1699, 688, 1120, 1120, 1120, 1120, 227, 581, 309, 1860, 804, 1
89, 1120, 133, 1120, 1120, 1120, 1120, 1120, 1373, 688, 992, 1095, 1120, 1120, 1120, 1120, 227, 1860, 1120, 1120, 1120,
666, 531, 1521, 493, 1689, 1120, 133, 1120, 1120, 1120, 1120, 1120, 1373, 688, 1120, 1120, 1120, 1120, 227, 1095, 1120,
1120, 1120, 1120, 1120, 1120, 1120, 1860, 804, 1689, 1120, 133, 1120, 1120, 1120, 1120, 688, 1536, 309, 1120, 1120, 1120, 112
, 227, 1860, 1036, 1120, 1120, 1120, 1120, 1120, 1450, 1120, 133, 1120, 1120, 1120, 1120, 1607, 953, 853, 1064, 401, 18
5, 1120, 133, 1120, 1120, 1120, 1120, 1790, 1120, 804, 354, 1741, 1662, 797, 277, 880, 1120, 1120, 354, 1722, 538, 1353
 1302, 1452, 1760, 1699, 1120, 1171, 953, 853, 1571, 1064, 1663, 1893, 756, 1699, 953, 1353, 1452, 1895, 614, 645, 1467
 877, 1788, 822, 953, 1353, 1452, 986, 953, 853, 1517, 74, 37, 1819, 1334, 681, 11, 1699, 1666, 1788, 176, 321, 74, 166
, 984, 1359, 953, 1353, 1452, 877, 617, 500, 74, 1452, 1893, 947, 1766, 771, 1893, 708, 1584, 1860, 1666, 176, 825, 166
, 374, 1171, 1099, 379, 1893, 238, 1452, 802, 1182, 571, 1162, 477, 227, 1405, 681, 1353, 688, 1452, 1191, 1493, 1852,
636, 1893, 945, 1663, 1120, 1048, 1699, 328, 1507, 71, 322, 1452, 1760, 1699, 1120, 1171, 53, 1741, 1162, 953, 853, 298
 775, 666, 1895, 681, 1353, 1330, 1895, 614, 877, 1666, 176, 869, 1893, 965, 997, 810, 1893, 825, 1162, 1676, 530, 825,
560, 74, 1182, 1859, 1699, 162, 1398, 688, 1191, 961, 1264, 298, 775, 74, 1741, 1398, 810, 1893, 1409, 825, 75, 1585, 1
40, 992, 666, 260, 1264, 1676, 530]

(base) C:\Users\alvar\Documents\GitHub\AprendizajeAutomatico\Prácticas\Práctica 6>
```