

ROC analysis with Support Vectors Machines

October 22, 2015

1 ROC Analysis for Multiclass SVMs

El analisis ROC es util para ilustrar el rendimiento de un clasificador generalmente binario (este ejemplo es la extencion para el caso multiclase). La curva se hace graficando la tasa de verdaderos positivos (i.e. TRP, recall, sensibilidad) contra la tasa de falsos positivos (i.e. FRP, especificidad, fallout) segn los diferentes ajustes de umbral.

1.1 True Positive Rate and False Positive Rate

```
In [19]: from sklearn.feature_extraction.text import TfidfVectorizer
         from sklearn.cross_validation import train_test_split, cross_val_score
```

```
tfidf_vect= TfidfVectorizer(use_idf=True, smooth_idf=True, sublinear_tf=False, ngram_range=(2,3))
import pandas as pd
```

```
df = pd.read_csv('/Users/user/Tesis/escritorio_anterior/MICAI/todos.csv',
                 header=0, sep=',', names=['id', 'content', 'label']).fillna(0)
df
```

```
Out[19]:
```

	id \
0	AEG_Electrolux_60840_Lavamat__Opinion_1506705
1	AEG_Electrolux_62610_Lavamat__Opinion_2000923
2	AEG_Electrolux_L14800VI__Opinion_2005396
3	AEG_Electrolux_L6227FL__Opinion_2140710
4	AEG_Electrolux_L62280FL__Opinion_2151025
5	AEG_Electrolux_L62642VI__Opinion_1995158
6	AEG_Electrolux_L62642VI__Opinion_1995791
7	AEG_Electrolux_L70470FL__Opinion_2085167
8	AEG_Electrolux_L74650__Opinion_1856967
9	AEG_Electrolux_L85275XFL__Opinion_2138100
10	AEG_Electrolux_L_14950_A__Opinion_2103767
11	AEG_Electrolux_L_47185__Opinion_1492496
12	AEG_Electrolux_L_47185__Opinion_1818711
13	AEG_Electrolux_L_47186__Opinion_1370192
14	AEG_Electrolux_L_47330_A__Opinion_2087353
15	AEG_Electrolux_L_5214__Opinion_1437024
16	AEG_Electrolux_L_62825__Opinion_1469992
17	AEG_Electrolux_L_70270_FL__Opinion_2090317
18	AEG_Electrolux_L_70270_FL__Opinion_2143750
19	AEG_Electrolux_L_70270_FL__Opinion_2150873
20	AEG_Electrolux_L_70270_FL__Opinion_2152723
21	AEG_Electrolux_L_70810__Opinion_1263244

22 AEG_Electrolux.L.70810..Opinion.1277946
 23 AEG_Electrolux.L.70810..Opinion.1284085
 24 AEG_Electrolux.L.70810..Opinion.1372777
 25 AEG_Electrolux.L.70810..Opinion.1848780
 26 AEG_Electrolux.L.70850..Opinion.1448444
 27 AEG_Electrolux.L.70850..Opinion.2007327
 28 AEG_Electrolux.L.725500..Opinion.1791610
 29 AEG_Electrolux.L.725500..Opinion.1794746

... ..
 2568 Zanussi_ZWH_6101_P..Opinion.1999819
 2569 Zanussi_ZWH_6101_P..Opinion.2001127
 2570 Zanussi_ZWH_6101_P..Opinion.2001296
 2571 Zanussi_ZWH_6101_P..Opinion.2001662
 2572 Zanussi_ZWH_6101_P..Opinion.2006735
 2573 Zanussi_ZWH_6101_P..Opinion.2007532
 2574 Zanussi_ZWH_6101_P..Opinion.2009115
 2575 Zanussi_ZWH_6101_P..Opinion.2026016
 2576 Zanussi_ZWH_6101_P..Opinion.2027055
 2577 Zanussi_ZWH_6102_PS..Opinion.2138338
 2578 Zanussi_ZWH_6120..Opinion.1981065
 2579 Zanussi_ZWH_6120..Opinion.2136981
 2580 Zanussi_ZWH_6125..Opinion.1701351
 2581 Zanussi_ZWH_7120_P..Opinion.2005676
 2582 Zanussi_ZWH_7120_P..Opinion.2132623
 2583 Zanussi_ZWH_7125_X..Opinion.2083788
 2584 Zanussi_ZWH_7125_X..Opinion.2127841
 2585 Zanussi_ZWK_5100..Opinion.1526854
 2586 Zanussi_ZWK_6120..Opinion.1613313
 2587 Zanussi_ZWN_5105..Opinion.2122850
 2588 Zanussi_ZWN_6105..Opinion.2004500
 2589 Zanussi_ZWN_6120..Opinion.2084720
 2590 Zanussi_ZWN_7100L..Opinion.2093055
 2591 Zanussi_ZWN_7120_L..Opinion.2131960
 2592 Zanussi_ZWN_7125..Opinion.2093504
 2593 Zanussi_ZWQ_5100..Opinion.2138972
 2594 Zanussi_ZWQ_5109..Opinion.2094551
 2595 Zanussi_ZWQ_598..Opinion.2004693
 2596 Zanussi_ZWT_260..Opinion.1091527
 2597 Zanussi_ZWT_260..Opinion.1157262

	content	label
0	programa especial ropa sensible planchado f\xe...	5
1	carga superior punto medio color blanco carga ...	4
2	carga superior forma frontal programa majo efi...	5
3	modelo electroluxmuy fcil altamente satisfact...	4
4	clasificacin energtico sonido bajo conversac...	4
5	ficha tecnica programa especial prenda delicad...	5
6	habitacin principal control electrnico clasi...	4
7	clasificacin energtico fase anterior lavado ...	4
8	capacidad grande entender mucho muy simple muy...	3
9	temperatura medio ciclo medio factor determina...	5
10	empresa alemn agua elctrico plancha elctric...	4
11	lavadora nuevo tipo energtico carga igual ten...	4
12	carga lateral carga superior mas caro no durar...	1

13	eta lavador carga superior ropa listo planchar...	4
14	lavadora normal persona principal acero inoxidable...	4
15	detalle tecnico punto central carga mximo cat...	5
16	quedar muy funcionar correctamente demasiado g...	4
17	mancha difcil ropa delicado lavadora normal c...	4
18	mvil directo horno pirolitico frigorifico ame...	5
19	mvil directo horno pirolitico frigorifico ame...	5
20	echar nada seguir as\x3\xad mucho menor muy c...	4
21	uso electronico lavadora viejo influir mucho e...	4
22	tiempo remanente programacin electrnico cont...	5
23	manejo diario ubicar comodamente bastante buen...	5
24	capacidad generoso aspecto exterior anchura ex...	3
25	mancha gris esla lavador uso suavizante ropa b...	3
26	acero inoxidable principio atractivo familia n...	4
27	tamao medio parte superior color azul color b...	5
28	manera sincero clasificacin energtico tiempo...	4
29	producto similarmuy bueno muy accesible muy fa...	4
...
2568	aspecto general uso diario clasificacin energ...	4
2569	clasificacin energtico control electrnico s...	4
2570	consumo energtico ahorro energtico medir ms...	4
2571	pronombre neutro lavadora antiguo color blanco...	3
2572	lavadora nuevo tener ya poco ruidoso muy contento	4
2573	cliente contento mvil directo esforzar mucho	5
2574	dimensin alto sistema aclarador eficacia ener...	5
2575	cliente contento mvil directo esforzar mucho	5
2576	cliente contento mvil directo esforzar mucho	5
2577	lavado super mas seco lavar bien salir tan acl...	3
2578	reducir bastante poner bastante tener muy bast...	4
2579	carga amplio lavadora viejo carga frontal ocup...	4
2580	parte superior servicio oficial servicio tecni...	2
2581	proceso actual broma jaja carga mximo sistema...	4
2582	eficiencia energtico cuanto principal lavador...	3
2583	lavado diario pantalla digital sistema antihue...	4
2584	mvil directo horno pirolitico frigorifico ame...	5
2585	piso atiguo clasificacion energetica muy bueno ...	4
2586	carga superior carga frontal carga frontal lav...	5
2587	prenda grande clasificacin energtico clasifi...	4
2588	lavadora fantstico display digital parte supe...	4
2589	salir ms hacer ya durar tanto tener hoy muy a...	3
2590	prenda grande clasificacin energtico clasifi...	5
2591	null	5
2592	corporacin sueco uso domestico placa vitrocer...	4
2593	euro aprox carga superior control estndar lav...	4
2594	lavadora fiable clasificacin energtico sper...	4
2595	mas grande display digital muy fcil no estrope...	2
2596	ao aprox cosa raro utilizar mucho salir muy i...	2
2597	lavadora normal lavadora pequeno pensar muy ce...	5

[2598 rows x 3 columns]

In []:

```
In [20]: X = tfidf_vect.fit_transform(df['content'].values)
y = df['label'].values
```

```

In [21]: # Binarize the output
         from sklearn.preprocessing import label_binarize

         y = label_binarize(y, classes=[1,2,3,4,5])
         n_classes = y.shape[1]
         random_state = 0

In [22]: # shuffle and split training and test sets
         X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=0)

         # Learn to predict each class against the other
         #from sklearn.ensemble import AdaBoostClassifier

         from sklearn.svm import SVC
         from sklearn.multiclass import OneVsRestClassifier

         classifier = OneVsRestClassifier(SVC(kernel = 'linear'))
         y_score = classifier.fit(X_train, y_train).decision_function(X_test)

In [23]: # Compute ROC curve and ROC area for each class
         from sklearn.metrics import roc_curve, auc

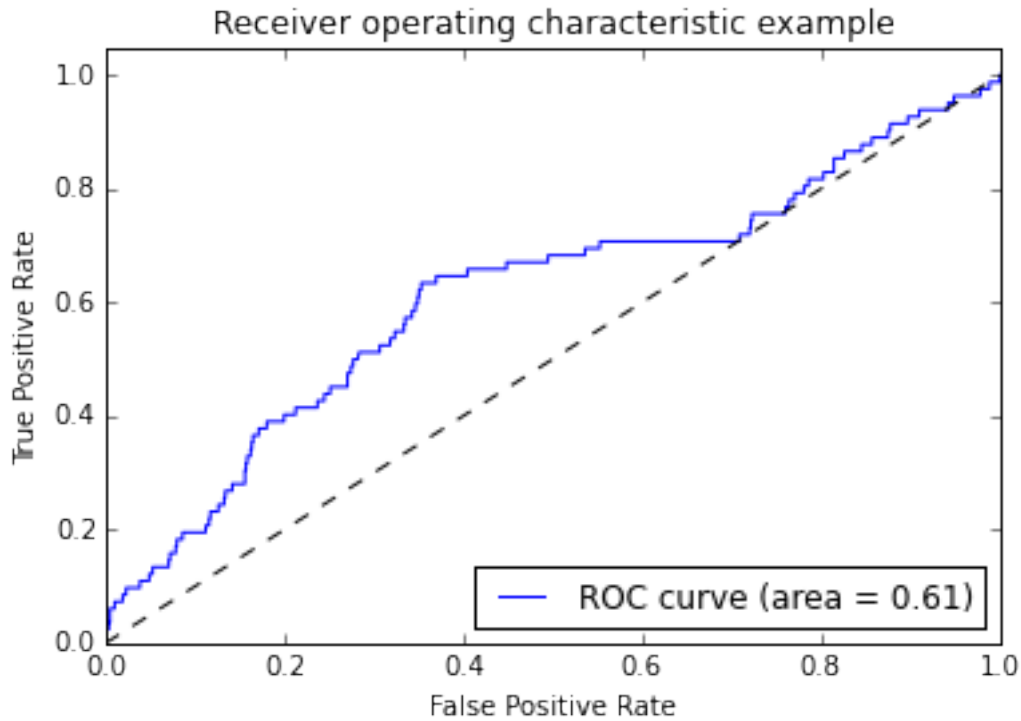
         fpr = dict()
         tpr = dict()
         roc_auc = dict()
         for i in range(n_classes):
             fpr[i], tpr[i], _ = roc_curve(y_test[:, i], y_score[:, i])
             roc_auc[i] = auc(fpr[i], tpr[i])

         # Compute micro-average ROC curve and ROC area
         fpr["micro"], tpr["micro"], _ = roc_curve(y_test.ravel(), y_score.ravel())
         roc_auc["micro"] = auc(fpr["micro"], tpr["micro"])

         %matplotlib inline
         import matplotlib.pyplot as plt

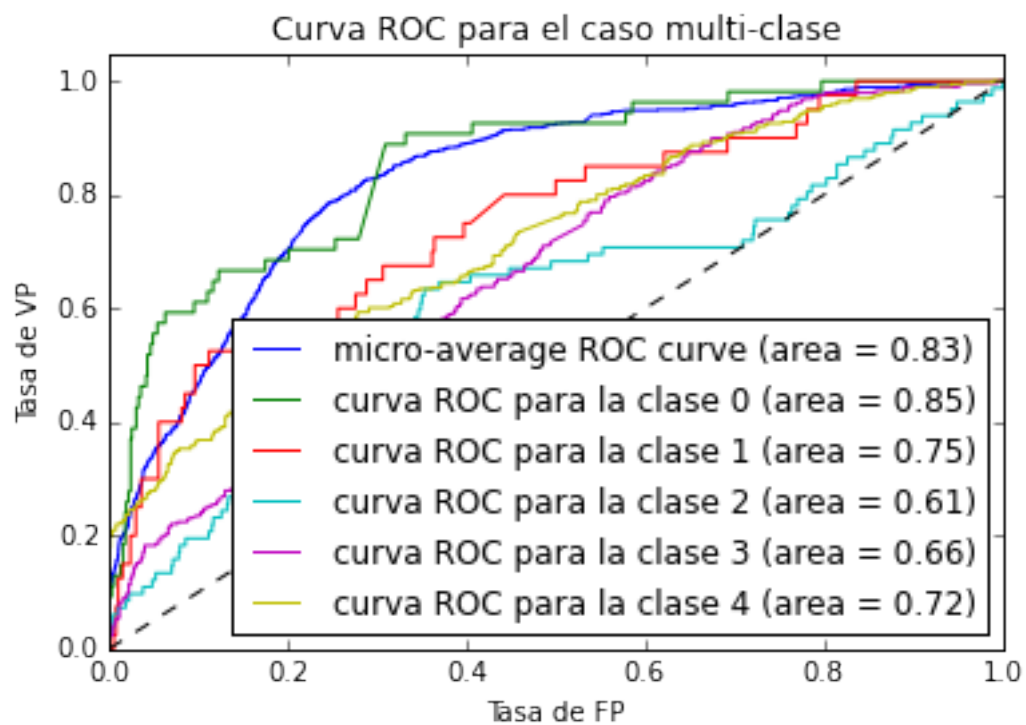
         # Plot of a ROC curve for a specific class
         plt.figure()
         plt.plot(fpr[2], tpr[2], label='ROC curve (area = %0.2f)' % roc_auc[2])
         plt.plot([0, 1], [0, 1], 'k--')
         plt.xlim([0.0, 1.0])
         plt.ylim([0.0, 1.05])
         plt.xlabel('False Positive Rate')
         plt.ylabel('True Positive Rate')
         plt.title('Receiver operating characteristic example')
         plt.legend(loc="lower right")
         plt.show()

```



```
In [28]: # Plot ROC curve
plt.figure()
plt.plot(fpr["micro"], tpr["micro"],
         label='micro-average ROC curve (area = {0:0.2f})'
         ''.format(roc_auc["micro"]))
for i in range(n_classes):
    plt.plot(fpr[i], tpr[i], label='curva ROC para la clase {0} (area = {1:0.2f})'
             ''.format(i, roc_auc[i]))

plt.plot([0, 1], [0, 1], 'k--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('Tasa de FP')
plt.ylabel('Tasa de VP')
plt.title('Curva ROC para el caso multi-clase')
plt.legend(loc="lower right")
plt.show()
```



In []:

In []: