



Evaluación de un modelo de clasificación

Ejemplos en WEKA

1. Calidad de vinos

El detalle de las columnas que tiene el dataset es:

acidez fija	densidad
acidez volátil	PH
ácido cítrico	sulfatos
azúcar residual	alcohol
cloruros	calidad
libre dióxido de azufre	
dióxido de azufre total	

Se busca determinar, qué características son las que contribuyen a una mejor calidad en el vino blanco. La calidad se especifica a partir de 5 clases: **3, 4, 5, 6, 7 y 8**.

Vamos a revisar algunos de los puntos sobre evaluación. Para esto, vamos a ejecutar una clasificación, utilizando resustitución y un árbol J48. Como se puede observar, se genera un modelo de clasificación muy grande, vamos a centrarnos en los aspectos de evaluación:

=== Run information ===

```
Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    calidad_vino
Instances:    1599
Attributes:   12
              acidez fija
              acidez volátil
              ácido cítrico
              residual sugar
              azúcar residual
              libre dióxido de azufre
              dióxido de azufre total
              densidad
              PH
              sulfatos
              alcohol
              calidad
Test mode:    evaluate on training data
```



```
=== Evaluation on training set ===
```

```
Time taken to test model on training data: 0.01 seconds
```

```
=== Summary ===
```

```
Correctly Classified Instances      1455           90.9944 %
Incorrectly Classified Instances    144            9.0056 %
Kappa statistic                    0.8584
Mean absolute error                 0.0465
Root mean squared error             0.1525
Relative absolute error             21.6818 %
Root relative squared error         46.5894 %
Total Number of Instances          1599
```

Podemos observar lo siguiente:

- Exactitud (correctos/total): $1455/1599 = 90.9944\%$
- Error de clasificación ($1 - \text{exactitud}$, incorrectos/total): $144/1599 = 9.0056\%$
- Concordancia: 85.84% (casi perfecta)
- Error medio absoluto (es una medida que se usa para medir qué tan cerca se encuentra la predicción del resultado real): 0.0465
- Error cuadrático medio (mide el promedio de los errores al cuadrado, es decir, la diferencia entre el estimador y lo que se estima. Amplifica y castiga severamente grandes errores, INDICA QUÉ TAN BUENO ES NUESTRO MODELO): 0.1525
- Error relativo absoluto (el error relativo porcentual refleja mejor la gravedad o no gravedad del error que se está cometiendo, error absoluto dividido entre el valor verdadero): 21.68%
- Raíz cuadrada del error relativo al cuadrado: (nos preocupamos si estuviera por arriba de 100%)

Vamos al detalle de las estadísticas:

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.931	0.050	0.932	0.931	0.932	0.881	0.983	0.974	q5
	0.937	0.075	0.893	0.937	0.914	0.856	0.977	0.963	q6
	0.869	0.009	0.930	0.869	0.899	0.886	0.992	0.955	q7
	0.660	0.005	0.814	0.660	0.729	0.725	0.992	0.817	q4
	0.667	0.003	0.750	0.667	0.706	0.704	0.996	0.729	q8
	0.300	0.001	0.750	0.300	0.429	0.472	0.997	0.591	q3
Weighted Avg.	0.910	0.053	0.909	0.910	0.908	0.862	0.982	0.957	

- **TP rate** (tasa de verdaderos positivos – Sensibilidad – recall): Porción de casos etiquetados con C1 que en realidad eran C1.
- **FP Rate** (Tasa de falsos positivos – Falsas alarmas): Porción de casos etiquetados como C2 que en realidad eran C1.
- **Precisión** (valor predictivo positivo): Probabilidad de un caso se etiquete con C1 que en realidad le corresponde C1.
- **Recall** (recuerdo, memoria) = TP Rate



- **F-Measure** (medida de precisión que tiene una prueba, medida armónica de la precisión y recall):
- **MCC** (coeficiente de correlación Matthews): es en esencia un coeficiente de correlación entre las clasificaciones binarias observadas y predichas; devuelve un valor entre -1 y +1. Un coeficiente de 1 representa una predicción perfecta, 0 no es mejor que la predicción aleatoria y -1 indica total desacuerdo entre la predicción y la observación.

El coeficiente de correlación Matthews se utiliza en la máquina de aprendizaje como una **medida de la calidad de las (dos clases) clasificaciones binarias**. Se calcula de la siguiente forma:

$$\frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Cuando se enfrentan con el problema del desequilibrio de clases, la precisión es una métrica incorrecto utilizar . Por lo general, hay dos candidatos como métricas:

- **ROC Area:** dan una idea de cómo los clasificadores están realizando su tarea en general. Dan el mismo resultado independientemente de lo que las probabilidades son de clase, es decir, que consideran igualmente las clases positivos y negativos:

La precisión se mide por el área bajo la curva ROC. Un área de 1 representa una prueba perfecta; un área de 0,5 representa una prueba de valor. Una guía general para la clasificación de la exactitud de una prueba de diagnóstico es el sistema tradicional punto académica:

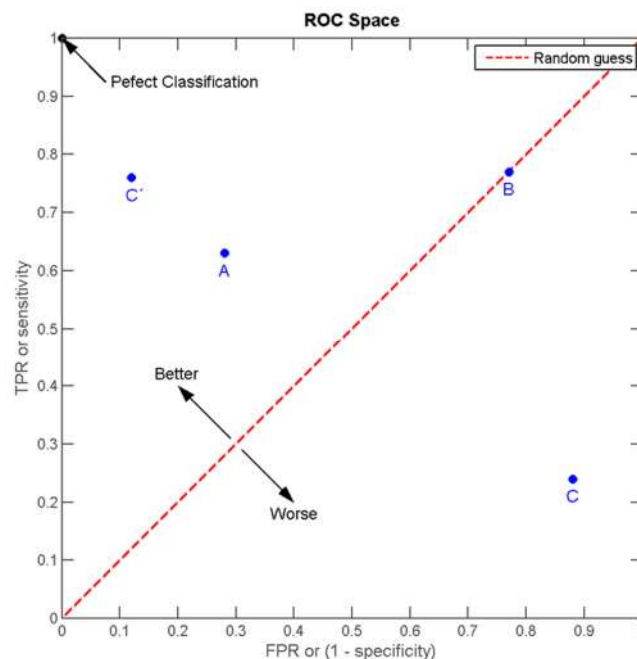
0.90 a 1 = excelente (A)

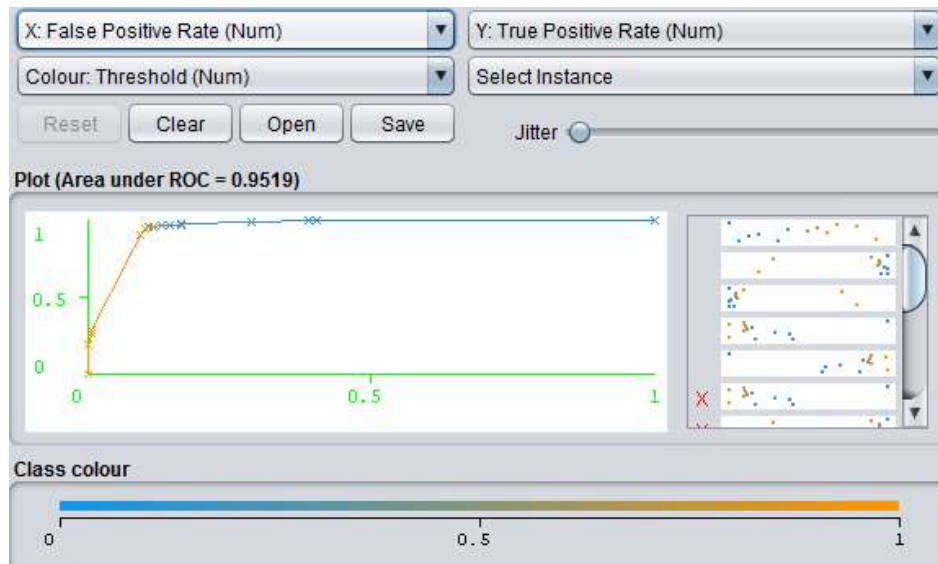
Desde 0.80 hasta 0.90 = bueno (B)

0.70-0.80 = Regular (C)

0.60 a 0.70 = deficiente (D)

0.50-0.60 = fallan (F)





En la curva ROC, el objetivo es tener un modelo esté en la esquina superior izquierda, que está recibiendo básicamente sin falsos positivos - un clasificador perfecto.

- **PCR Area (gráficas de precisión – recall):** PCR es más útil si estamos interesados sólo en la forma en que el clasificador se está comportando de una clase. Imagine que trata de clasificar a los pacientes como enfermos o sanos. En este caso, usted no está interesado en el número de predicciones saludables son correctas, desea predecir correctamente todos los casos enfermos y no omitir ninguna.

El objetivo es tener un modelo esté en la esquina superior derecha, que es básicamente recibiendo sólo los verdaderos positivos sin falsos positivos y falsos negativos - un clasificador perfecto.

