

Ajuste, Evaluación y Sintonía de Modelos



Universidad
Católica del
Uruguay

Evaluación de los modelos

- Métodos para determinar el desempeño de un modelo de clasificación:
 - Matriz de confusión
 - Curva ROC (receiver operator characteristic)
 - Lift Chart

Matriz de confusión

- Consideremos un caso de clasificación binario Y/N (yes/no)
 - Si el valor real es Y y el predicho Y, estamos ante un “positivo verdadero” (true positive, o TP)
 - Si el valor real es N y el predicho Y, estamos ante un “falso positivo” (false positive, o FP)
 - Si el valor real es N y el predicho N, estamos ante un “negativo verdadero” (true negative, o TN)
 - Si el valor real es Y y el predicho N, estamos ante un “falso negativo” (false negative, o FN)

Matriz de confusión

		Actual Class(Observation)	
		Y	N
Predicted Class (Expectation)	Y	TP (true positive) Correct result	FP (false positive) Unexpected result
	N	FN (false negative) Missing result	TN (true negative) Correct absence of result

- En un caso de clasificación perfecta, FN y FP serían 0
- El mismo razonamiento puede extenderse para múltiples clases

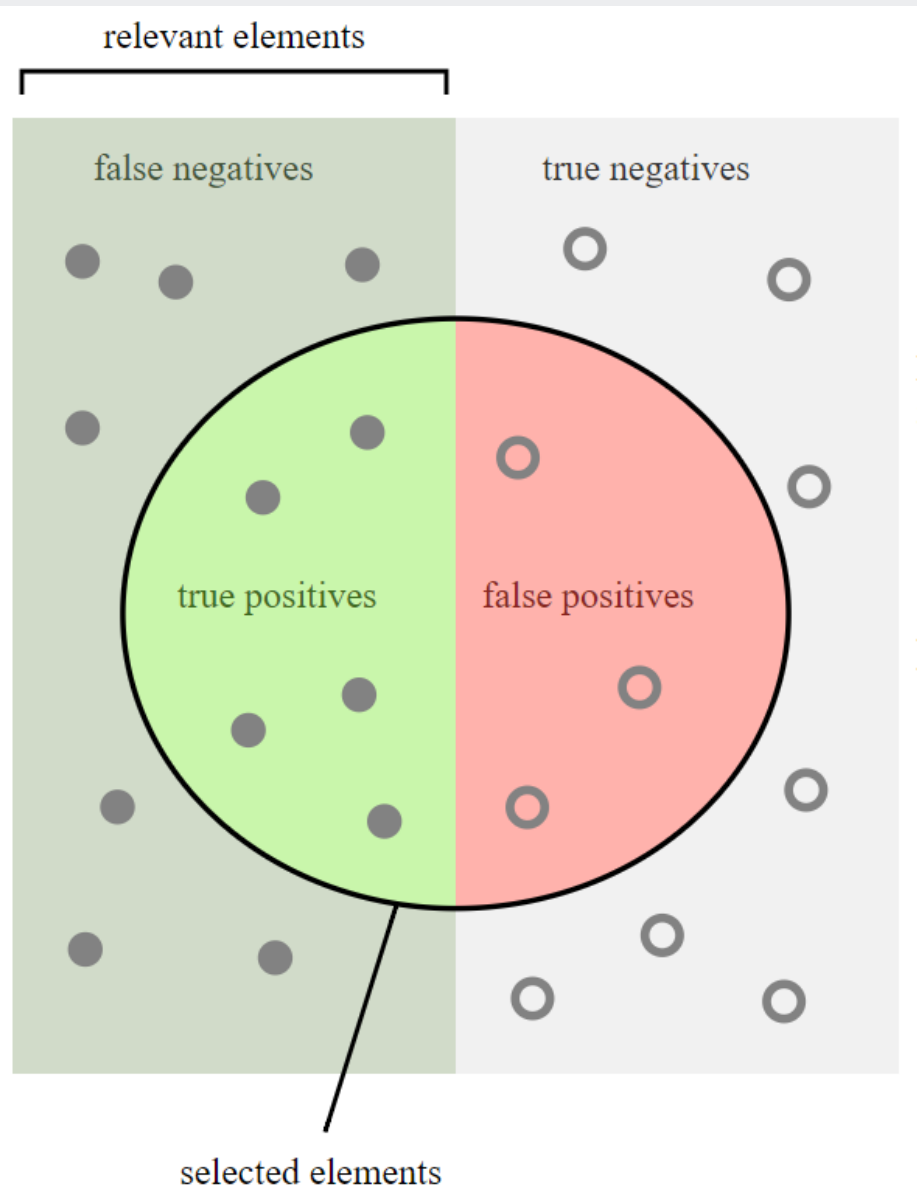
Otras métricas

- Sensibilidad: si consideramos la clase de interés a “Y”, mide la habilidad del clasificador de predecir correctamente todos los ejemplos Y del dataset
- $TP/(TP + FN)$
- El caso ideal es $FN=0$
- En algunos problemas puede no ser suficiente:
 - Si la clase indica si una transacción es válida, y ocurren transacciones fraudulentas con 0.1% de probabilidad, predecir todo como válido generaría un 99.9% de sensibilidad, lo cual no es útil

Otras métricas

- Especificidad: mide la habilidad del clasificador de rechazar todos los casos que deben ser rechazados (es decir, sin falsos positivos)
- $TN / (TN + FP)$

Otras métricas: Precision y Recall



How many selected
items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant
items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

Otras métricas

- Precision: de todos los elementos que predecimos que son de la clase de interés, qué proporción lo hacemos correctamente?
 - $TP/(TP+FP)$
- Recall: de todos los elementos que son verdaderamente de interés, qué proporción logramos predecir?
 - $TP/(TP+FN)$

Otras métricas

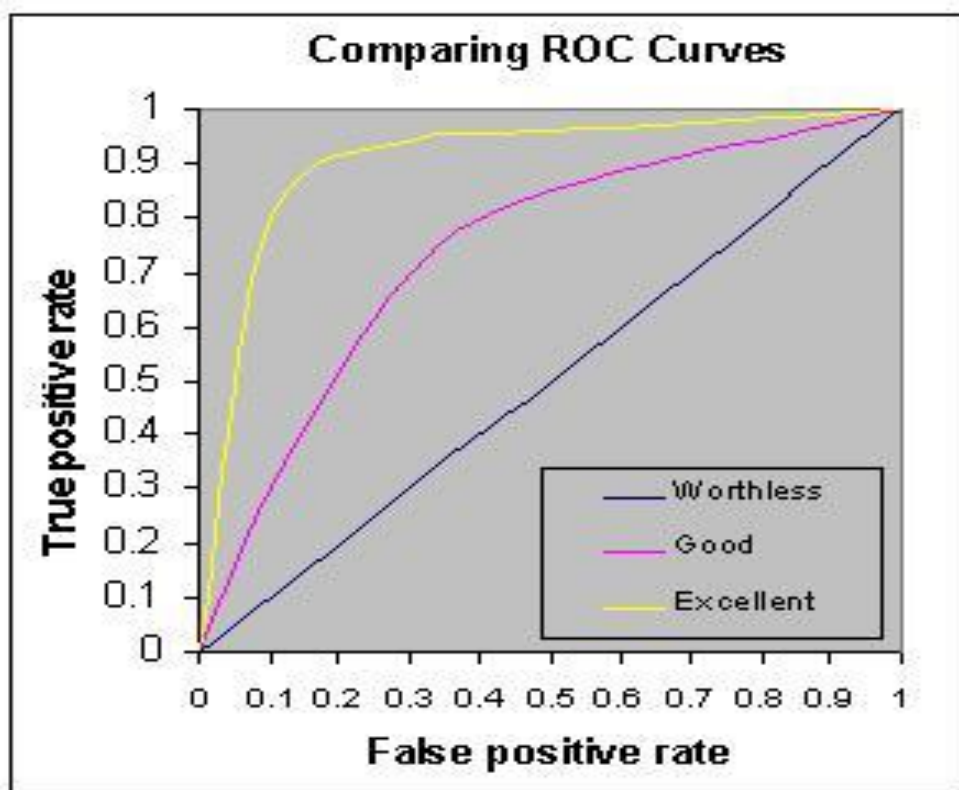
- Accuracy: mide la capacidad del clasificador de predecir correctamente las clases
 - 100% de accuracy implica $FN = FP = 0$
 - $(TP+TN)/(TP+FP+TN+FN)$
- Error: $(1 - \text{accuracy})$

Curva ROC

- Un clasificador puede tener un valor de accuracy alto pero valores bajos de precision y/o recall
- Ej: Sea un modelo de detección de fraude, donde indicamos con la clase positiva la ocurrencia de un fraude. Una medida de recall bajo (de todos los fraudes predecimos una proporción baja) es no deseable.
- ¿Podemos sacrificar un poco el accuracy para aumentar efectivamente el recall?
- La curva ROC (receiver operator characteristic) permite medir esto

Curva ROC

- En una curva ROC graficamos la fracción de verdaderos positivos que van apareciendo en los datos (eje vertical) vs. la fracción de falsos positivos (eje horizontal)



- La recta (0,0) (1,1) simboliza un modelo con un rendimiento similar a predicciones aleatorias (misma tasa de falsos positivos y verdaderos positivos) - AUROC=0.5
- Curvas por encima indican un mejor rendimiento del modelo

Caso de Estudio: “Affinity Based Marketing”

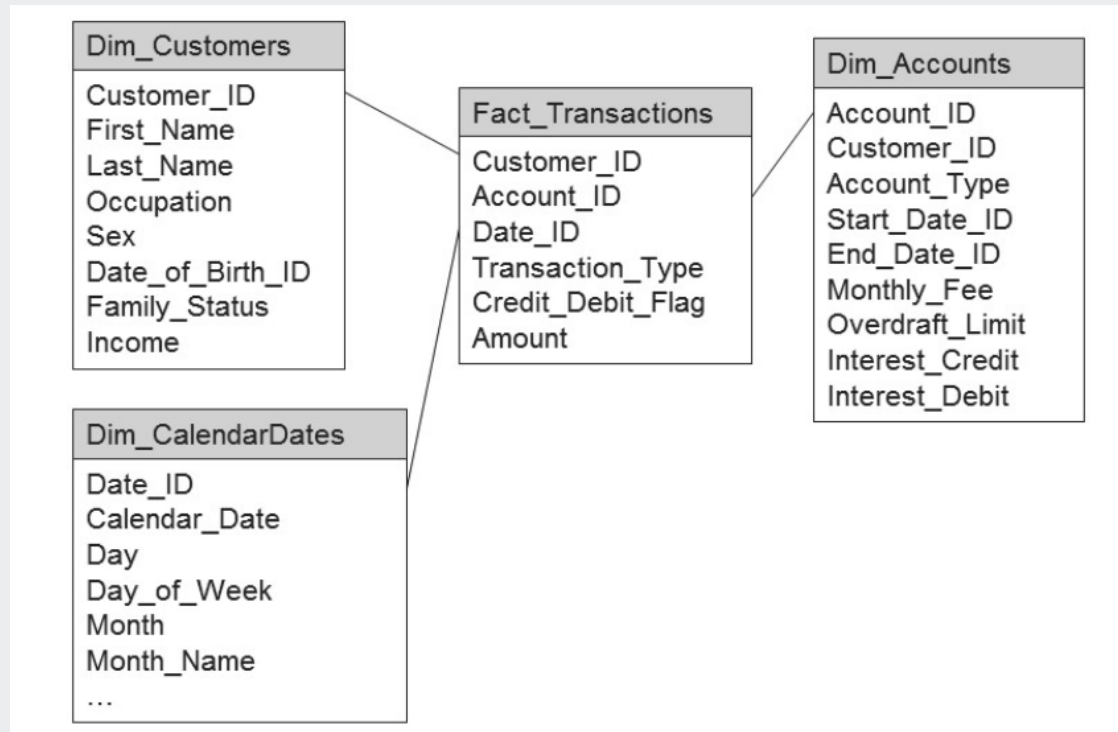


Caso de estudio

- Descarga el documento adjunto a la tarea
- Crea una carpeta en tu repositorio RM y pon allí los datasets provistos.
- PASO 1: comprensión del problema y del contexto (secciones 7.1 y 7.2)
 - Documentar!
 - ¿características de las cuentas?
 - ¿transacciones? ¿tipos?
 - ¿clientes? ¿qué información hay?
 - ¿qué productos pueden comprar los clientes?
 - *Comportamiento* del cliente

PASO 2: Comprensión de los datos

- Documentar tablas, relaciones, campos (sección 7.3).



- Analizar, comprender y documentar las simplificaciones que se han realizado (ver doc.)

PASO 3: Preparación de los datos (1)



- Sección 7.4
- Los algoritmos de ML habitualmente utilizan una tabla única integrada y “curada”
- Integración de los datasets de entrada
- Generación de un dataset único para aplicar a tareas de ML
- proceso: *01 CreateMiningTable*
 - Ver que se utilizan las 4 tablas como entradas, y se produce una de salida “RawMiningTable”
 - Mostrar el orden de ejecución
 - Tipos de datos...
 - ¿cuáles son los “sub-canales” del proceso existentes, y qué hacen?
 - Analizar y documentar el resultado...

PASO 3: Preparación de los datos (2)



- Sub-stream “customer data”
 - Documentar las operaciones realizadas y su objetivo, así como los operadores utilizados y los parámetros disponibles
- Sub-stream “account data”
 - ¿Cuál es el objetivo de este proceso?
 - Documentar las operaciones realizadas y su objetivo, así como los operadores utilizados y los parámetros disponibles
- Sub-stream: “transaction data”
 - Ídem
- Fase final de integración de los sub-streams

PASO 3: Preparación de los datos (3)



- Preparación de los datos para minería
- Proceso: *“02 PrepareDataForMining”*
 - Utiliza la salida del proceso anterior
 - Aplica una cadena sencilla de operadores para dejar los datos listos para el modelado
 - Observar el uso del operador *“Set Role”*
 - ¿qué atributo se usa como ID? ¿cómo funciona?
- *“Data cleaning”* –
 - valores faltantes: ¿qué se hace? ¿Cuáles serían las alternativas posibles?
 - Atributos inútiles...
- Discretización – analizar y documentar qué se hace y por qué

PASO 3: Preparación de los datos (4)



- Revisión de proceso generado y estructura del dataset integrado resultante
- Explicar al otro equipo, preguntas y respuestas

PASO 4: Modelado y Evaluación



- Sección 7.5
- Generacion de modelo, separación train-test
- Cross validation, desbalance de clases
- Evaluación simple

PASO 5: Modelado y Evaluación

- Probar diferentes modelos (buscar mejores resultados)
- Valores de confianza, curvas ROC y gráficos LIFT

PASO 6: despliegue

- Sección 7.6
- Hallazgos, conclusiones

Preguntas...

