



FCFM

Maestría en Ciencias de Datos

Aprendizaje Automático

Tarea 8: Métricas de Desempeño

Edwin Martín Romero Silva

3 de noviembre de 2023

Resumen

1. Parte 1 Métrica de error a una BD vista en clase
 - Base de datos elegida
 - Error Cuadrado Medio
 - Desempeño
2. Parte 2 Métricas de error en mi problema
 - Métricas Adecuadas
 - Curva ROC y AUC
 - KS: kolmogorov-smirnov
 - Aplicación de la métrica desempeño
3. Conclusiones

1. Parte 1: Métricas de error a una BD vista en clase

1.1. Base de datos elegida

Para esta parte de la tarea escogí la base de datos candy.csv, la cual contiene las siguientes variables:

- Chocolate: bandera 0/1
- Fruity: bandera 0/1
- Caramel: bandera 0/1
- Peanutyalmondy: bandera 0/1
- Nougat: bandera 0/1
- Crispedricewafer: bandera 0/1
- Hard: bandera 0/1
- Bar: bandera 0/1
- Pluribus: bandera 0/1
- Sugarpercent: Continua
- Pricepercent: Continua
- Winpercent: Continua

Escogí la variable winpercent como variable objetivo para un problema de regresión. Hice el entrenamiento un Gradient Boosting Regressor con los siguientes hiperparámetros:

```
GB = GradientBoostingRegressor(n_estimators = 125,  
                                learning_rate = 0.05,  
                                max_depth     = 5,  
                                max_features  = 0.7,  
                                subsample     = 1)
```

Figura 1: Hiperparámetros de XGB en Python.

1.2. Error Cuadrado Medio

Utilicé la métrica del Error Cuadrado Medio para medir el desempeño de este modelo, me parece adecuada ya que hablamos de un problema de regresión.

Consiste en calcular el promedio de los cuadrados de las diferencias entre las predicciones del modelo y los valores reales en un conjunto de datos de prueba.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

El MSE mide cómo se dispersan los errores cuadráticos entre las predicciones y los valores reales. Cuanto más bajo sea el valor del MSE, mejor será el rendimiento del modelo. Recomendando sacar la raíz cuadrada a este valor de MSE para una mejor interpretación.

1.3. Desempeño

El modelo XGB obtiene el siguiente desempeño con la métrica mencionada:

- MSE: 111.4679
- sqrt(MSE): 10.5578

2. Parte 2 Métricas de Desempeño en mi problema

2.1. Métricas adecuadas

Mi problema es de clasificación binaria, por lo que necesito métricas de desempeño que me permitan medir que tan bueno es mi modelo para separar los ‘Buenos’ de los ‘Malos’. 2 métricas muy útiles para este tipo de problemas son:

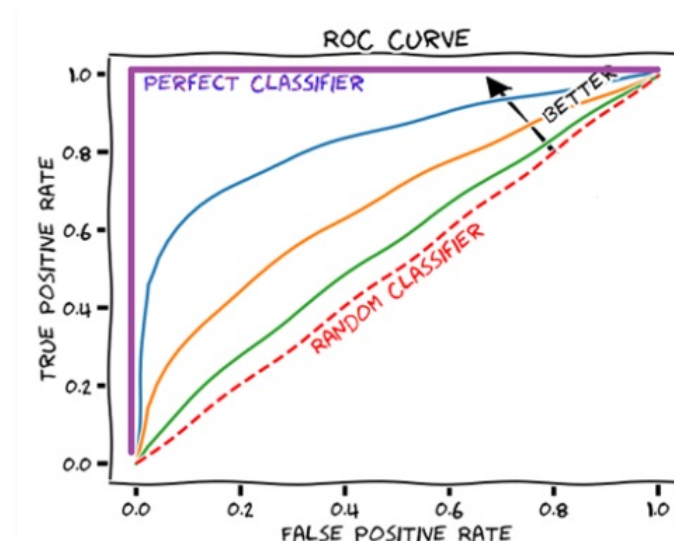
2.1.1. ROC y AUC

Primero, el ROC es una representación gráfica que se utiliza comúnmente en problemas de clasificación para evaluar el rendimiento de un modelo, particularmente en modelos binarios. La curva ROC muestra la relación entre dos tasas:

Tasa de Verdaderos Positivos (TPR): Mide la capacidad del modelo para identificar instancias positivas. Esta es la proporción de positivos reales que el modelo clasifica correctamente como positivos.

Tasa de Falsos Positivos (FPR): Mide la tasa de errores en los que el modelo clasifica incorrectamente instancias negativas como positivas. Esta es la proporción de negativos reales que el modelo clasifica incorrectamente como positivos.

La curva ROC traza la TPR en el eje vertical y la FPR en el eje horizontal a medida que se ajusta el umbral de decisión del modelo.



El AUC, es el área debajo de la curva ROC. Cuanto mayor sea el AUC, mejor será el rendimiento del modelo. El AUC varía de 0 a 1, donde:

- AUC menor a 0.5 indica que el modelo es peor que una decisión aleatoria.
- AUC igual a 0.5 indica que el modelo tiene un rendimiento similar al aleatorio.
- AUC mayor a 0.5 indica que el modelo tiene cierta capacidad de discriminación y es mejor que una decisión aleatoria.
- AUC igual a 1 indica un clasificador perfecto que clasifica todas las instancias positivas antes que las negativas.

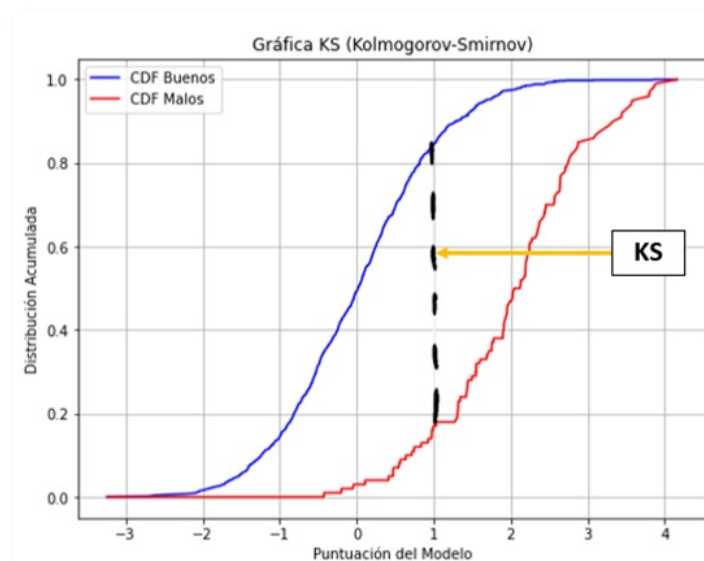
2.1.2. Índice de Kolmogorov – Smirnov

Un modelo de clasificación binaria nos entrega la probabilidad de que un caso sea marcado como ‘Malo’ (en el caso de mi problema).

n	Probabilidad Malo
1	0.25
2	0.10
3	0.37
4	0.19
...	...

Cuadro 1: Probabilidad de Impago.

El KS, representa la máxima diferencia absoluta entre la distribución acumulada de las probabilidades de los clientes Buenos y la distribución acumulada de las probabilidades de los clientes Malos.



2.2. Aplicación de métricas de desempeño a mi problema

Utilicé la métrica AUC de la siguiente forma:

Entrené 675 modelos de Gradient Boosting Classifier, cada uno con una distinta combinación de hiperparámetros y a cada uno de ellos le calculé el AUC tanto en train como en test. De forma que escogeré el modelo con mejor desempeño en test, según la métrica del AUC.

OP_ESTIMATORS	OP_SUBSAMPLE	OP_MAX_DEPTH	OP_LEARNING_RATE	OP_LOSS	OP_TRAIN_SAMPLE	AUC_TRAIN	AUC_TEST
311	150	0.75	25	0.20	log_loss	0.8	0.788037
202	125	0.90	15	0.20	log_loss	0.7	0.804444
569	250	0.75	20	0.15	log_loss	0.8	0.791882
290	150	0.75	15	0.15	log_loss	0.8	0.792830
626	250	0.90	25	0.20	log_loss	0.8	0.794898
...
229	125	1.00	9	0.20	log_loss	0.7	0.764316
513	200	1.00	15	0.15	log_loss	0.6	0.775037
404	150	1.00	25	0.25	log_loss	0.8	0.762936
90	100	1.00	9	0.15	log_loss	0.6	0.761672
227	125	1.00	9	0.15	log_loss	0.8	0.741959

675 rows x 8 columns

El mejor modelo encontrado tiene un AUC de 0.62, lo cual es mejor que tomar decisiones aleatorias, pero sigue siendo pobre ya que se acerca mucho a 0.5.

3. Conclusión

Las métricas de error o de desempeño nos permiten medir que tan bueno es nuestro modelo para cumplir su objetivo y de esta forma tomar decisiones. Yo las utilicé para escoger la mejor combinación de hiperparámetros, por ejemplo.

Debemos escoger la métrica mas adecuada para nuestro problema, ya sea de regresión o de clasificación. En mi caso yo utilicé Error Cuadrado Promedio para regresión y Área Bajo la Curva para clasificación binaria.