

Machine Learning Aplicado

Medidas de Evaluación

Ingeniería Civil Informática
Escuela de Ingeniería Informática
Universidad de Valparaíso

Profesor: Aaron Ponce Sandoval
correo: aaron.ponce@uv.cl

Una vez que se entrena un problema de ML supervisado en un conjunto de datos históricos, se prueba mediante el uso de datos que se retienen del proceso de entrenamiento. De esta forma, puede compararse las predicciones del modelo entrenado con los valores reales.

Las métricas de evaluación

Las métricas de evaluación son medidas que se utilizan para evaluar el rendimiento de un modelo de machine learning. Estas métricas proporcionan una forma objetiva de medir el modelo. Permiten observar que tan bien un modelo está realizando predicciones en nuevos datos.

- ❑ Permiten a los científicos de datos y a los ingenieros de machine learning comparar diferentes modelos
- ❑ Permite identificar posibles problemas con un modelo, como overfitting o underfitting.



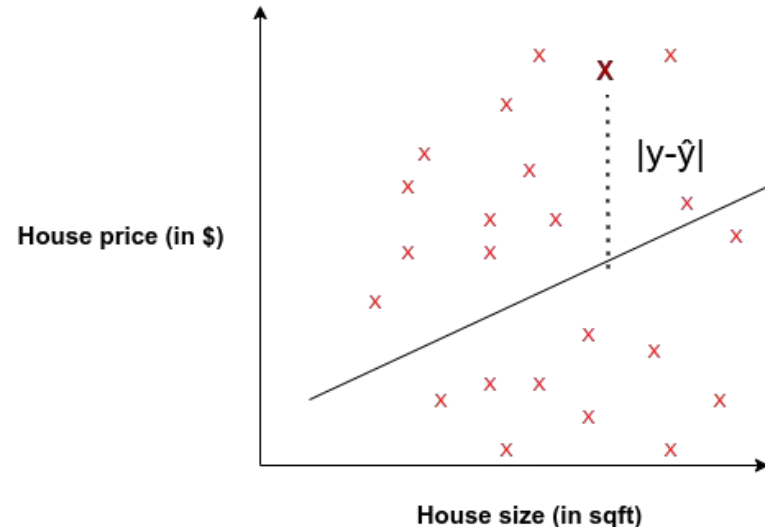
En los modelos de regresión, la salida es una variable continua y para evaluar la precisión de las predicciones necesitamos una métrica que permita calcular la distancia entre los valores predichos y los valores reales.

- ❑ Mean Absolute Error (MAE),
- ❑ Mean Squared Error (MSE),
- ❑ Error cuadrático medio de la raíz (RMSE),
- ❑ R^2 (R-Squared).

Mean Absolute Error (MSE),

El error absoluto medio es una medida que calcula el promedio de las diferencias entre los valores reales y los valores predichos por el modelo. Se puede representar matemáticamente como la suma de las diferencias absolutas entre los valores reales y los valores predichos, dividida por el número total de observaciones.

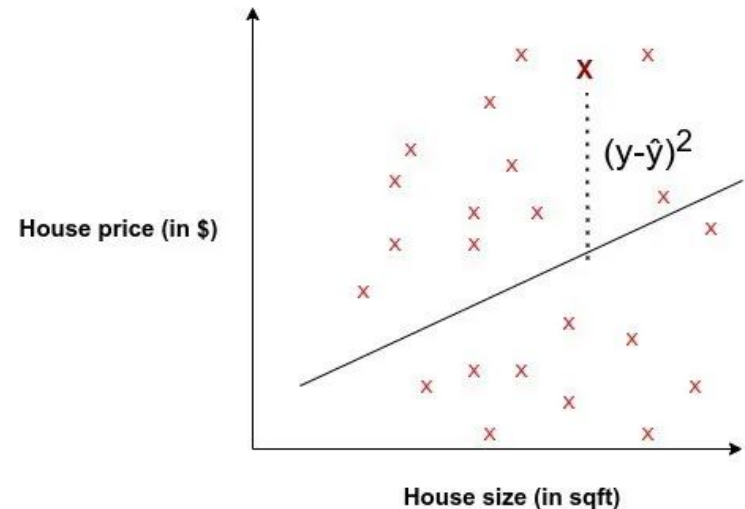
$$MAE = \frac{1}{N} \sum_{j=1}^N |y_j - \hat{y}_j|$$



Mean Squared Error (MSE)

El método más comúnmente utilizado para medir el rendimiento de los modelos de regresión es el error cuadrático medio, que calcula la media de las diferencias al cuadrado entre los valores predichos por el modelo y los valores reales. Esta métrica es muy popular porque proporciona una medida cuantitativa de la precisión del modelo y es fácil de entender e interpretar.

$$MSE = \frac{1}{N} \sum_{j=1}^N (y_j - \hat{y}_j)^2$$



Root Mean Squared Error (RMSE)

El error cuadrático medio corresponde a la raíz cuadrada del promedio de la diferencia cuadrática entre el valor objetivo y el valor predicho por el modelo de regresión. Básicamente, $\sqrt{\text{MSE}}$

$$RMSE = \sqrt{\frac{1}{N} \sum_{j=1}^N (y_j - \hat{y}_j)^2}$$

Debido a que los modelos de clasificación producen resultados discretos, se requiere una métrica que pueda comparar y evaluar el rendimiento del modelo en la clasificación. Las métricas de clasificación son herramientas que se utilizan para evaluar la calidad de la predicción de un modelo. Sin embargo, cada una de ellas utiliza un enfoque diferente para evaluar el desempeño del modelo de clasificación.

- ☐ Accuracy
- ☐ Confusion Matrix (no es métrica)
- ☐ Precision and Recall
- ☐ F1-score
- ☐ AU-ROC

Confusion Matrix

Confusion Matrix es una visualización tabular de las etiquetas verdad frente a las predicciones del modelo. Cada fila de la matriz de confusión representa las instancias de una clase predicha y cada columna representa las instancias de una clase real.

| | | Predicción | |
|-------------|-----------|---------------------------|---------------------------|
| | | Positivos | Negativos |
| Observación | Positivos | Verdaderos Positivos (VP) | Falsos Negativos (FN) |
| | Negativos | Falsos Positivos (FP) | Verdaderos Negativos (VN) |

- **VP** es la cantidad de *positivos* que fueron *clasificados correctamente* como positivos por el modelo.
- **VN** es la cantidad de *negativos* que fueron *clasificados correctamente* como negativos por el modelo.
- **FN** es la cantidad de *positivos* que fueron *clasificados incorrectamente* como negativos.
- **FP** es la cantidad de *negativos* que fueron *clasificados incorrectamente* como positivos.

Verdadero Positivo (TP): Se refiere al número de predicciones donde el clasificador predice correctamente la clase positiva como positiva.

True Negative (TN): Se refiere al número de predicciones donde el clasificador predice correctamente la clase negativa como negativa.

Falso Positivo (FP): Se refiere al número de predicciones donde el clasificador predice incorrectamente la clase negativa como positiva.

Falso Negativo (FN): Se refiere al número de predicciones donde el clasificador predice incorrectamente la clase positiva como negativa

Verdadero Positivo (TP): Se refiere al número de predicciones donde el clasificador predice correctamente la clase positiva como positiva.

True Negative (TN): Se refiere al número de predicciones donde el clasificador predice correctamente la clase negativa como negativa.

Falso Positivo (FP): Se refiere al número de predicciones donde el clasificador predice incorrectamente la clase negativa como positiva.

Falso Negativo (FN): Se refiere al número de predicciones donde el clasificador predice incorrectamente la clase positiva como negativa

Ejemplo Binario

A continuación se exponen los resultados obtenidos de un clasificador **binario** que clasifica si una paciente tiene o no tiene cáncer.

| | | Predicted | |
|--------------|---------------------|------------|---------------------|
| | | Has Cancer | Doesn't Have Cancer |
| Ground Truth | Has Cancer | TP | FP |
| | Doesn't Have Cancer | FN | TN |

Ejemplo Multi-Class

En este caso el modelo predice si la factura de un cliente se pagará a tiempo, tarde o muy tarde. Por ejemplo, de 101 facturas de prueba, 50 se pagan a tiempo, 35 se pagan con retraso y 16 se pagan con mucho retraso. En este caso, un modelo puede producir una matriz de confusión similar a la siguiente ilustración.

| P R E D I C C I O N | Valores Verdaderos | | | |
|--|--------------------|-------------|-------|--------------|
| | | A Tiempo | Tarde | Muy Tarde |
| | A Tiempo | 40 | 5 | 3 |
| | Tarde | 7 | 25 | 5 |
| | Muy Tarde | 3 | 5 | 8 |

Matriz A Tiempo

TP: 40

FN: $5+3 = 8$

FP: $7+3=10$

TN: $25+5+5+8=38$

Matriz Tarde

TP: 25

FN: $7+5 = 12$

FP: $5+3=8$

TN: $40+3+3+8=54$

Matriz Muy Tarde

TP: 8

FN: $3+5 = 8$

FP: $3+5=8$

TN: $40+7+5+25=77$

Accuracy (Exactitud)

La precisión de la clasificación es una métrica que se utiliza para determinar el porcentaje de predicciones correctas en comparación con el número total de predicciones realizadas. Es calculada dividiendo el número de predicciones correctas por el número total de predicciones y luego multiplicando por 100 para obtener el porcentaje.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

Mide la proporción de verdaderos positivos (TP) en relación con todos los resultados positivos pronosticados por un modelo. Es decir, es la relación entre los verdaderos positivos y la suma de los verdaderos positivos y falsos positivos (FP).

- ❑ Una puntuación de precisión cercana a 1 significa que su modelo no pasó por alto ningún positivo verdadero
- ❑ Una puntuación de precisión baja (< 0.5) significa que su clasificador tiene una gran cantidad de **falsos positivos**

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall (Sensibilidad)

Recall mide la capacidad de un modelo para encontrar todos los casos positivos. En otras palabras, el recall es la proporción de verdaderos positivos que fueron identificados correctamente sobre el total de casos positivos en los datos.

- ❑ El recuerdo hacia 1 significa que su modelo no pasó por alto ningún positivo verdadero
- ❑ Una puntuación de recuperación baja ($<0,5$) significa que su clasificador tiene una gran cantidad de **falsos negativos**

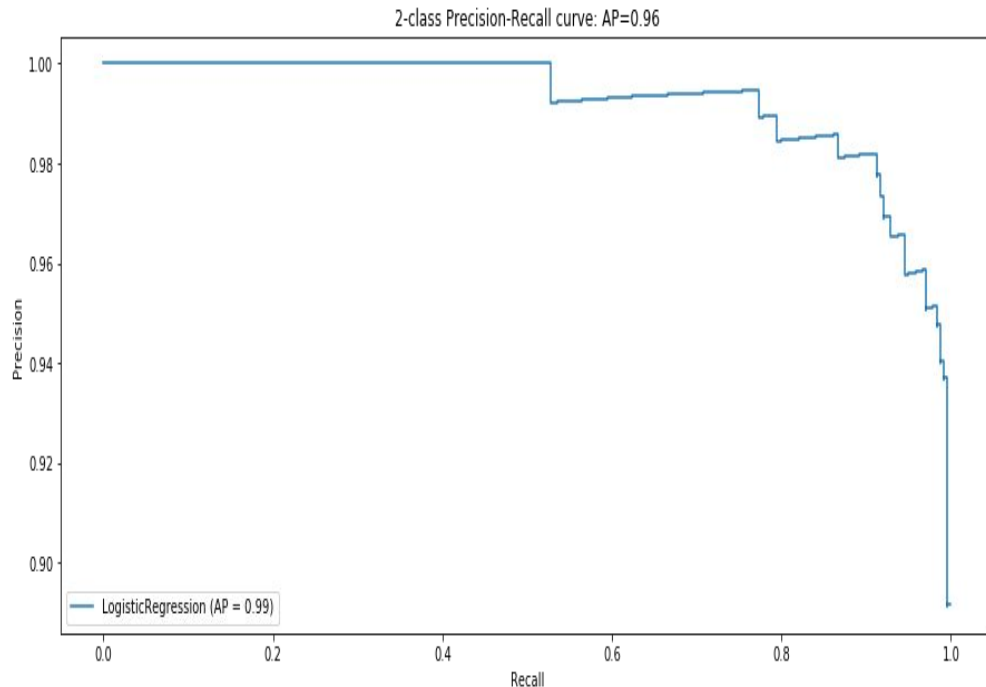
$$\text{Recall} = \frac{TP}{TP + FN}$$

Compensación de recuperación de precisión

Para mejorar su modelo, puede mejorar la precisión o la recuperación, ¡pero no ambas!

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$



La métrica de puntuación F1 utiliza una combinación de precisión y recuperación. Esta puntuación F1 es la media armónica de las dos.

- ❑ La métrica F1-score alta simboliza una alta precisión y una alta recuperación.
- ❑ La métrica F1-score alta presenta un buen equilibrio entre precisión y recuperación y da buenos resultados en problemas de clasificación desequilibrada.
- ❑ Un F1 bajo, no dice muchos del problema (¿baja precisión o poca recuperación?)

$$F_1 = \frac{2}{\frac{1}{precision} + \frac{1}{recall}}$$