

# Hoja de referencia VIP: Consejos y trucos sobre Aprendizaje Automático

Afshine AMIDI y Shervine AMIDI

6 de octubre de 2018

Traducido por David Jiménez Paredes y Fernando Díaz. Revisado por Gustavo Velasco-Hernández y Alonso Melgar-Lopez.

## Métricas para clasificación

En el contexto de una clasificación binaria, estas son las principales métricas que son importantes seguir para evaluar el rendimiento del modelo.

□ **Matriz de confusión** – La matriz de confusión (en inglés, *Confusion matrix*) se utiliza para tener una visión más completa al evaluar el rendimiento de un modelo. Se define de la siguiente manera:

		Clase predicha	
		+	-
Clase real	+	<b>TP</b> True Positives	<b>FN</b> False Negatives Type II error
	-	<b>FP</b> False Positives Type I error	<b>TN</b> True Negatives

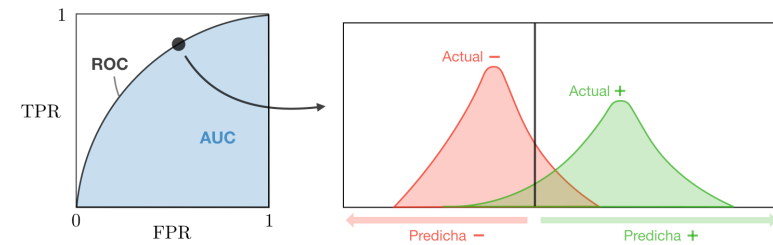
□ **Métricas principales** – Las siguientes métricas se utilizan comúnmente para evaluar el rendimiento de los modelos de clasificación:

Métrica	Fórmula	Interpretación
Exactitud	$\frac{TP + TN}{TP + TN + FP + FN}$	Rendimiento general del modelo
Precisión	$\frac{TP}{TP + FP}$	Que tan precisas son las predicciones positivas
Exhaustividad Sensibilidad	$\frac{TP}{TP + FN}$	Cobertura de la muestra positiva real
Especificidad	$\frac{TN}{TN + FP}$	Cobertura de la muestra negativa real
F1 score	$\frac{2TP}{2TP + FP + FN}$	Métrica híbrida útil para clases desbalanceadas

□ **ROC** – La curva Característica Operativa del Receptor (en inglés, *Receiver Operating Curve*), también conocida como ROC, es una representación gráfica de la sensibilidad frente a la especificidad según se varía el umbral. Estas métricas se resumen en la siguiente tabla:

Métrica	Fórmula	Interpretación
True Positive Rate TPR	$\frac{TP}{TP + FN}$	Exhaustividad, sensibilidad
False Positive Rate FPR	$\frac{FP}{TN + FP}$	1-especificidad

□ **AUC** – El área bajo la curva Característica Operativa del Receptor, también conocida como AUC o AUROC (en inglés, *Area Under the Receiving Operating Curve*), es el área debajo del ROC, como se muestra en la siguiente figura:



## Métricas de regresión

□ **Métricas básicas** – Dado un modelo de regresión  $f$ , las siguientes métricas se usan comúnmente para evaluar el rendimiento del modelo:

Suma total de cuad.	Suma de cuad. explicada	Suma residual de cuad.
$SS_{\text{tot}} = \sum_{i=1}^m (y_i - \bar{y})^2$	$SS_{\text{reg}} = \sum_{i=1}^m (f(x_i) - \bar{y})^2$	$SS_{\text{res}} = \sum_{i=1}^m (y_i - f(x_i))^2$

□ **Coefficiente de determinación** – El coeficiente de determinación, a menudo indicado como  $R^2$  o  $r^2$ , proporciona una medida de lo bien que los resultados observados son replicados por el modelo y se define de la siguiente manera:

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

□ **Métricas principales** – Las siguientes métricas se utilizan comúnmente para evaluar el rendimiento de los modelos de regresión, teniendo en cuenta la cantidad de variables  $n$  que tienen en consideración:

Cp de Mallow	AIC	BIC	$R^2$ ajustado
$\frac{SS_{\text{res}} + 2(n+1)\hat{\sigma}^2}{m}$	$2[(n+2) - \log(L)]$	$\log(m)(n+2) - 2\log(L)$	$1 - \frac{(1 - R^2)(m-1)}{m-n-1}$

donde  $L$  es la probabilidad y  $\hat{\sigma}^2$  es una estimación de la varianza asociada con cada respuesta.

### Selección de modelo

□ **Vocabulario** – Al seleccionar un modelo, distinguimos 3 partes diferentes de los datos que tenemos de la siguiente manera:

Entrenamiento	Validación	Prueba
<ul style="list-style-type: none"> <li>- Modelo es entrenado</li> <li>- Generalmente el 80 % del conjunto de datos</li> </ul>	<ul style="list-style-type: none"> <li>- Modelo es evaluado</li> <li>- Generalmente 20 %</li> <li>- También llamado hold-out o conjunto de desarrollo</li> </ul>	<ul style="list-style-type: none"> <li>- Modelo da predicciones</li> <li>- Datos no vistos</li> </ul>

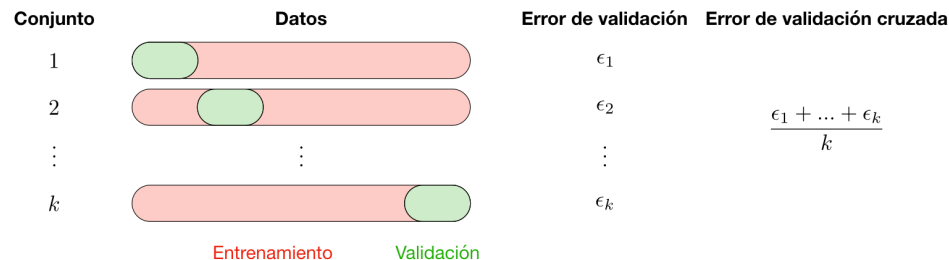
Una vez que se ha elegido el modelo, se entrena sobre todo el conjunto de datos y se testea sobre el conjunto de prueba no visto. Estos están representados en la figura a continuación:



□ **Validación cruzada** – La validación cruzada, también denominada CV (en inglés, *Cross validation*), es un método que se utiliza para seleccionar un modelo que no confíe demasiado en el conjunto de entrenamiento inicial. Los diferentes tipos se resumen en la tabla a continuación:

$k$ -fold	Leave- $p$ -out
<ul style="list-style-type: none"> <li>- Entrenamiento sobre los conjuntos <math>k-1</math> y evaluación en el restante</li> <li>- Generalmente <math>k = 5</math> o <math>10</math></li> </ul>	<ul style="list-style-type: none"> <li>- Entrenamiento en observaciones <math>n-p</math> y evaluación en los <math>p</math> restantes</li> <li>- El caso <math>p = 1</math> se llama <i>leave-one-out</i></li> </ul>

El método más comúnmente utilizado se denomina validación cruzada  $k$ -fold y divide los datos de entrenamiento en  $k$  conjuntos para validar el modelo sobre un conjunto mientras se entrena el modelo en los otros  $k-1$  conjuntos, todo esto  $k$  veces. El error luego se promedia sobre los  $k$  conjuntos y se denomina error de validación cruzada.



□ **Regularización** – El procedimiento de regularización tiene como objetivo evitar que el modelo se sobreajuste a los datos y, por lo tanto, resuelve los problemas de alta varianza. La siguiente tabla resume los diferentes tipos de técnicas de regularización comúnmente utilizadas:

LASSO	Ridge	Elastic Net
<ul style="list-style-type: none"> <li>- Reduce los coeficientes a 0</li> <li>- Bueno para la selección de variables</li> </ul>	Hace que los coeficientes sean más pequeños	Compensación entre la selección de variables y los coeficientes pequeños
$\dots + \lambda   \theta  _1$ $\lambda \in \mathbb{R}$	$\dots + \lambda   \theta  _2^2$ $\lambda \in \mathbb{R}$	$\dots + \lambda \left[ (1-\alpha)  \theta  _1 + \alpha  \theta  _2^2 \right]$ $\lambda \in \mathbb{R}, \alpha \in [0,1]$

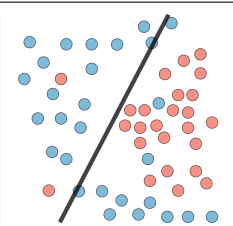
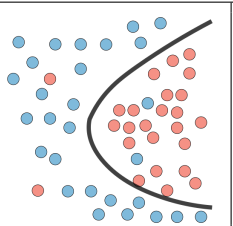
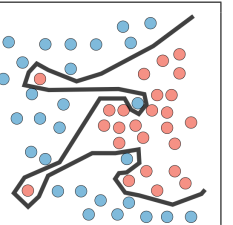
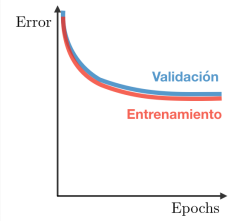
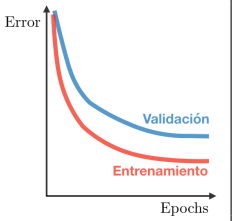
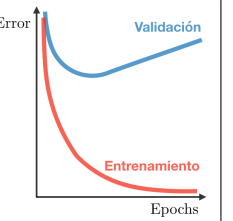
### Diagnóstico

□ **Sesgo** – El sesgo (en inglés, *Bias*) de un modelo es la diferencia entre la predicción esperada y el modelo correcto que tratamos de predecir para determinados puntos de datos.

□ **Varianza** – La varianza (en inglés, *Variance*) de un modelo es la variabilidad de la predicción del modelo para puntos de datos dados.

□ **Corrección de sesgo/varianza** – Cuanto más simple es el modelo, mayor es el sesgo, y cuanto más complejo es el modelo, mayor es la varianza.

	Underfitting	Just right	Overfitting
<b>Síntomas</b>	<ul style="list-style-type: none"> <li>- Error de entrenamiento alto</li> <li>- Error de entrenamiento cercano al error de prueba</li> <li>- Sesgo alto</li> </ul>	<ul style="list-style-type: none"> <li>- Error de entrenamiento légèrement inférieure à l'erreur de test</li> </ul>	<ul style="list-style-type: none"> <li>- Error de entrenamiento muy bajo</li> <li>- Error de entrenamiento mucho más bajo que el error de prueba</li> <li>- Varianza alta</li> </ul>
<b>Regresión</b>			

<b>Clasificación</b>			
<b>Deep Learning</b>			
<b>Soluciones</b>	<ul style="list-style-type: none"> <li>- Incrementar la complejidad del modelo</li> <li>- Agregar más funciones</li> <li>- Entrenar más tiempo</li> </ul>		<ul style="list-style-type: none"> <li>- Realizar la regularización</li> <li>- Obtener más datos</li> </ul>

□ **Análisis de errores** – El análisis de errores analiza la causa raíz de la diferencia de rendimiento entre los modelos actuales y perfectos.

□ **Análisis ablativo** – El análisis ablativo analiza la causa raíz de la diferencia en el rendimiento entre los modelos actuales y de referencia.