

Dicas e truques de aprendizado de máquina

Afshine AMIDI e Shervine AMIDI

13 de Outubro de 2018

Traduzido por Fernando Santos. Revisado por Leticia Portella e Gabriel Fonseca.

Métricas de classificação

Em um contexto de classificação binária, essas são as principais métricas que são importantes acompanhar para avaliar a desempenho do modelo.

□ **Matriz de confusão** – A matriz de confusão (*confusion matrix*) é usada para termos uma cenário mais completa quando estamos avaliando o desempenho de um modelo. Ela é definida conforme a seguir:

		Classe prevista	
		+	-
Classe real	+	TP True Positives	FN False Negatives Type II error
	-	FP False Positives Type I error	TN True Negatives

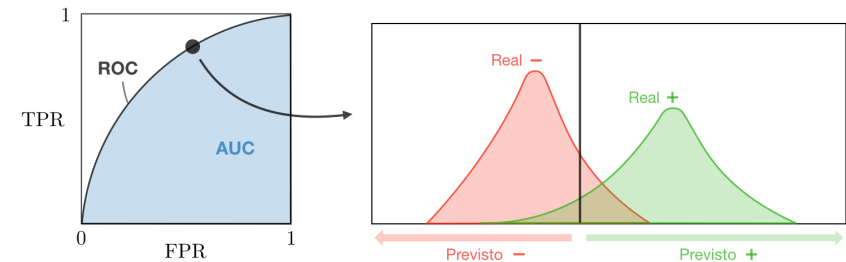
□ **Principais métricas** – As seguintes métricas são comumente usadas para avaliar o desempenho de modelos de classificação:

Métrica	Fórmula	Interpretação
Acurácia	$\frac{TP + TN}{TP + TN + FP + FN}$	Desempenho geral do modelo
Precisão	$\frac{TP}{TP + FP}$	Quão precisas são as predições positivas
Revocação Sensibilidade	$\frac{TP}{TP + FN}$	Cobertura da amostra positiva real
Specificity	$\frac{TN}{TN + FP}$	Cobertura da amostra negativa real
F1 score	$\frac{2TP}{2TP + FP + FN}$	Métrica híbrida útil para classes desequilibradas

□ **ROC** – A curva de operação do receptor, também chamada ROC (*Receiver Operating Characteristic*), é a área de TPR versus FPR variando o limiar. Essa métricas estão resumidas na tabela abaixo:

Métrica	Fórmula	Equivalente
True Positive Rate TPR	$\frac{TP}{TP + FN}$	Revocação, sensibilidade
False Positive Rate FPR	$\frac{FP}{TN + FP}$	1-specificity

□ **AUC** – A área sob a curva de operação de recebimento, também chamado AUC ou AUROC, é a área abaixo da ROC como mostrada na figura a seguir:



Métricas de regressão

□ **Métricas básicas** – Dado um modelo de regressão f , as seguintes métricas são geralmente utilizadas para avaliar o desempenho do modelo:

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

□ **Coefficiente de determinação** – O coeficiente de determinação, frequentemente escrito como R^2 ou r^2 , fornece uma medida de quão bem os resultados observados são replicados pelo modelo e é definido como se segue:

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

□ **Principais métricas** – As seguintes métricas são comumente utilizadas para avaliar o desempenho de modelos de regressão, levando em conta o número de variáveis n que eles consideram:

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

onde L é a probabilidade e $\hat{\sigma}^2$ é uma estimativa da variância associada com cada resposta.

Seleção de modelo

□ **Vocabulário** – Ao selecionar um modelo, nós consideramos 3 diferentes partes dos dados que possuímos conforme a seguir:

Conjunto de treino	Conjunto de validação	Conjunto de teste
<ul style="list-style-type: none"> - Modelo é treinado - Geralmente 80% do conjunto de dados 	<ul style="list-style-type: none"> - Modelo é avaliado - Geralmente 20% do conjunto de dados - Também chamado de hold-out 	<ul style="list-style-type: none"> - Modelo fornece previsões - Dados não vistos

Uma vez que o modelo é escolhido, ele é treinado no conjunto inteiro de dados e testado no conjunto de dados de testes não vistos. São representados na figura abaixo:



□ **Validação cruzada** – Validação cruzada, também chamada de CV (*Cross-Validation*), é um método utilizado para selecionar um modelo que não depende muito do conjunto de treinamento inicial. Os diferentes tipos estão resumidos na tabela abaixo:

<i>k</i> -fold	Leave- <i>p</i> -out
<ul style="list-style-type: none"> - Treino em $k - 1$ partes e teste sobre o restante - Geralmente $k = 5$ ou 10 	<ul style="list-style-type: none"> - Treino em $n - p$ observações e teste sobre p restantes - Caso $p = 1$ é chamado <i>leave-one-out</i>

O método mais comumente usado é chamado *k*-fold cross validation e divide os dados de treinamento em k partes enquanto treina o modelo nas outras $k - 1$ partes, todas estas em k vezes. O erro é então calculado sobre as k partes e é chamado erro de validação cruzada (*cross-validation error*).

Parte	Dados	Erro de validação	Erro de validação cruzada
1		ϵ_1	$\frac{\epsilon_1 + \dots + \epsilon_k}{k}$
2		ϵ_2	
\vdots	\vdots	\vdots	
k		ϵ_k	

□ **Regularização** – O procedimento de regularização (*regularization*) visa evitar que o modelo sobreajuste os dados e portanto lide com os problemas de alta variância. A tabela a seguir resume os diferentes tipos de técnicas de regularização comumente utilizadas:

LASSO	Ridge	Elastic Net
<ul style="list-style-type: none"> - Diminui coeficientes para 0 - Bom para seleção de variáveis 	Faz o coeficiente menor	Balanço entre seleção de variáveis e coeficientes pequenos
$\dots + \lambda \theta _1$ $\lambda \in \mathbb{R}$	$\dots + \lambda \theta _2^2$ $\lambda \in \mathbb{R}$	$\dots + \lambda \left[(1 - \alpha) \theta _1 + \alpha \theta _2^2 \right]$ $\lambda \in \mathbb{R}, \alpha \in [0, 1]$

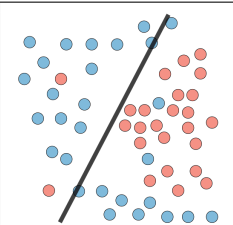
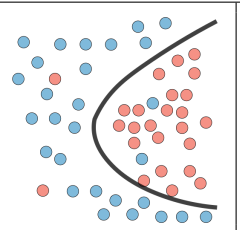
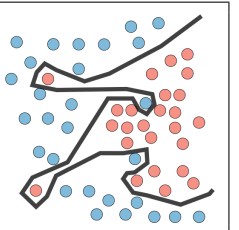
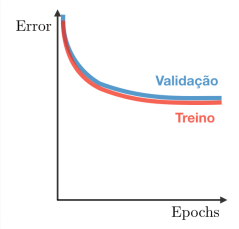
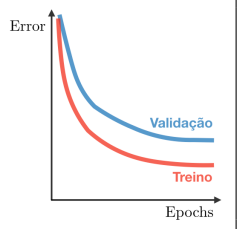
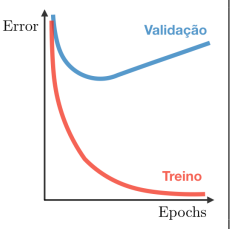
Diagnóstico

□ **Viés** – O viés (*bias*) de um modelo é a diferença entre a predição esperada e o modelo correto que nós tentamos prever para determinados pontos de dados.

□ **Variância** – A variância (*variance*) de um modelo é a variabilidade da previsão do modelo para determinados pontos de dados.

□ **Balanço viés/variância** – Quanto mais simples o modelo, maior o viés e, quanto mais complexo o modelo, maior a variância.

	Underfitting	Just right	Overfitting
Sintomas	<ul style="list-style-type: none"> - Erro de treinamento elevado - Erro de treinamento próximo ao erro de teste - Viés elevado 	<ul style="list-style-type: none"> - Erro de treinamento ligeiramente menor que erro de teste 	<ul style="list-style-type: none"> - Erro de treinamento muito baixo - Erro de treinamento muito menor que erro de teste - Alta variância
Regressão			

Classificação			
Deep Learning			
Remédios	<ul style="list-style-type: none"> - Modelo de complexificação - Adicionar mais recursos - Treinar mais 		<ul style="list-style-type: none"> - Executar a regularização - Obter mais dados

□ **Análise de erro** – Análise de erro (*error analysis*) é a análise da causa raiz da diferença no desempenho entre o modelo atual e o modelo perfeito.

□ **Análise ablativa** – Ablative analysis (*ablative analysis*) é a análise da causa raiz da diferença no desempenho entre o modelo atual e o modelo base.