

المقياس	المعادلة	التفسير
الضبط (accuracy)	$\frac{TP + TN}{TP + TN + FP + FN}$	الأداء العام للنموذج
Precision	$\frac{TP}{TP + FP}$	دقة التوقعات الإيجابية (positive)
Recall Sensitivity	$\frac{TP}{TP + FN}$	تغطية عينات التوقعات الإيجابية الفعلية
Specificity	$\frac{TN}{TN + FP}$	تغطية عينات التوقعات السلبية الفعلية
درجة F1	$\frac{2TP}{2TP + FP + FN}$	مقياس هجين مفيد للأصناف غير المتوازنة (unbalanced)

مرجع سريع لنصائح وحيل تعلم الآلة

افشين عميدى و شروين عميدى

١٤ ربيع الثاني، ١٤٤١

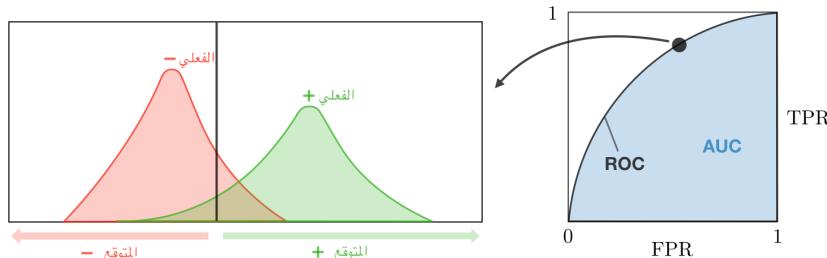
□ **منحنى دقة الأداء (ROC)** - منحنى دقة الأداء، ويطلق عليه ROC، هو رسم لمعدل التصنيفات الإيجابية الصحيحة (TPR) مقابل معدل التصنيفات الإيجابية الخاطئة (FPR) باستخدام قيم حد (threshold) متغيرة. هذه المقاييس ملخصة في الجدول التالي:

المقياس	المعادلة	مرادف
True Positive Rate TPR	$\frac{TP}{TP + FN}$	Recall, sensitivity
False Positive Rate FPR	$\frac{FP}{TN + FP}$	1-specificity

مقاييس التصنيف

في سياق التصنيف الثنائي، هذه المقاييس (metrics) المهمة التي يجدر مراقبتها من أجل تقييم أداء النموذج.

□ **المساحة تحت منحنى دقة الأداء المساحة تحت المنحنى (AUC)** - المساحة تحت منحنى دقة الأداء المساحة تحت المنحنى، ويطلق عليها AUC أو AUROC، هي المساحة تحت ROC كما هو موضح في الرسم التالية:



مقاييس الانحدار

□ **المقاييس الأساسية** - إذا كان لدينا نموذج الانحدار f ، فإن المقاييس التالية غالباً ما تستخدم لتقييم أداء النموذج:

□ **مصفوفة الدقة (confusion matrix)** - تستخدم مصفوفة الدقة لأخذ تصور شامل عند تقييم أداء النموذج. وهي تعرّف كالتالي:

		التصنيف المتوقع		
		-	+	
التصنيف الفعلي	+	FN False Negatives Type II error	TP True Positives	+
	-	TN True Negatives	FP False Positives Type I error	-

□ **المقاييس الأساسية** - المقاييس التالية تستخدم في العادة لتقييم أداء نماذج التصنيف:

Leave-p-out	k-fold
- التدريب على $n - p$ عينة والتقييم باستخدام الـ p عينات المتبقية - الحالة $p = 1$ يطلق عليها الإبقاء على واحد (leave-one-out)	- التدريب على $k - 1$ جزء والتقييم باستخدام الجزء الباقي - بشكل عام $k = 5$ أو 10

الطريقة الأكثر استخداماً يطلق عليها التحقق المتقاطع س جزء (k-fold)، ويتم فيها تقسيم البيانات إلى k جزء، بحيث يتم تدريب النموذج باستخدام $k - 1$ والتحقق باستخدام الجزء المتبقي، ويتم تكرار ذلك k مرة. يتم بعد ذلك حساب معدل الأخطاء في الأجزاء k ويسمى خطأ التحقق المتقاطع.

خطئ في التحقق	بيانات	جزء
ϵ_1		1
ϵ_2		2
\vdots	\vdots	\vdots
ϵ_k		k
$\frac{\epsilon_1 + \dots + \epsilon_k}{k}$	تحقق	تدريب

❑ **ضبط (Regularization)** - عملية الضبط تهدف إلى تفادي فرط التخصيص (overfit) للنموذج، وهو بذلك يتعامل مع مشاكل التباين العالي. الجدول التالي يلخص أنواع وطرق الضبط الأكثر استخداماً:

Elastic Net	Ridge	LASSO
المفاضلة بين اختيار المتغيرات والمعاملات الصغيرة	يجعل المعاملات أصغر	- يقلص المعاملات إلى ٠ - جيد لاختيار المتغيرات
$(1 - \alpha) \theta _1 + \alpha \theta _2^2 \leq 1$	$ \theta _2 \leq 1$	$ \theta _1 \leq 1$
$\dots + \lambda \left[(1 - \alpha) \theta _1 + \alpha \theta _2^2 \right]$ $\lambda \in \mathbb{R}, \alpha \in [0, 1]$	$\dots + \lambda \theta _2^2$ $\lambda \in \mathbb{R}$	$\dots + \lambda \theta _1$ $\lambda \in \mathbb{R}$

المجموع الكلي للمربعات	مجموع المربعات المُفسَّر	مجموع المربعات المتبقي
$SS_{\text{tot}} = \sum_{i=1}^m (y_i - \bar{y})^2$	$SS_{\text{reg}} = \sum_{i=1}^m (f(x_i) - \bar{y})^2$	$SS_{\text{res}} = \sum_{i=1}^m (y_i - f(x_i))^2$

❑ **معامل التحديد (Coefficient of determination)** - مُعامل التحديد، وغالباً يرمز له بـ R^2 أو r^2 ، يعطي قياس لمدى مطابقة النموذج للنتائج الملحوظة، ويعرف كما يلي:

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

❑ **المقاييس الرئيسية** - المقاييس التالية تستخدم غالباً لتقييم أداء نماذج الانحدار، وذلك بأن يتم الأخذ في الحسبان عدد المتغيرات n المستخدمة فيها:

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

حيث L هو الأرجحية، و $\hat{\sigma}^2$ تقدير التباين الخاص بكل نتيجة.

اختيار النموذج

❑ **مفردات** - عند اختيار النموذج، نفرق بين ٣ أجزاء من البيانات التي لدينا كالتالي:

مجموعة اختبار	مجموعة تحقق	مجموعة تدريب
- النموذج يعطي التوقعات - بيانات لم يسبق رؤيتها من قبل	- يتم تقييم النموذج - غالباً 20% من مجموعة البيانات - يطلق عليها كذلك المجموعة المُجنّبة أو مجموعة التطوير	- يتم تدريب النموذج - غالباً 80% من مجموعة البيانات

بمجرد اختيار النموذج، يتم تدريبه على مجموعة البيانات بالكامل ثم يتم اختباره على مجموعة اختبار لم يسبق رؤيتها من قبل. كما هو موضح في الشكل التالي:



❑ **التحقق المتقاطع (Cross-validation)** - التحقق المتقاطع، وكذلك يختصر بـ CV، هو طريقة تستخدم لاختيار نموذج بحيث لا يعتمد بشكل كبير على مجموعة بيانات التدريب المبدئية. أنواع التحقق المتقاطع المختلفة ملخصة في الجدول التالي:

التشخيصات

□ تحليل الخطأ - تحليل الخطأ هو تحليل السبب الرئيسي للفرق في الأداء بين النماذج الحالية والنماذج المثالية.

□ تحليل استقصائي (Ablative analysis) - التحليل الاستقصائي هو تحليل السبب الرئيسي للفرق في الأداء بين النماذج الحالية والنماذج المبدئية (baseline).

□ الانحياز (Bias) - الانحياز للنموذج هو الفرق بين التنبؤ المتوقع والنموذج الحقيقي الذي نحاول تنبؤه للبيانات المعطاة.

□ التباين (Variance) - تباين النموذج هو مقدار التغير في تنبؤ النموذج لنقاط البيانات المعطاة.

□ موازنة الانحياز/التباين (Bias/variance tradeoff) - كلما زادت بساطة النموذج، زاد الانحياز، وكلما زاد تعقيد النموذج، زاد التباين.

Overfitting	Just right	Underfitting	
- خطأ التدريب منخفض جداً - خطأ التدريب أقل بكثير من خطأ الاختبار - تباين عالي	- خطأ التدريب أقل بقليل من خطأ الاختبار	- خطأ التدريب عالي - خطأ التدريب قريب من خطأ الاختبار - انحياز عالي	الأعراض
			توضيح الانحدار
			توضيح التصنيف
			توضيح التعلم العميق
- إجراء الضبط (regularization) - الحصول على المزيد من البيانات		- زيادة تعقيد النموذج - إضافة المزيد من الخصائص - تدريب لمدة أطول	العلاجات الممكنة