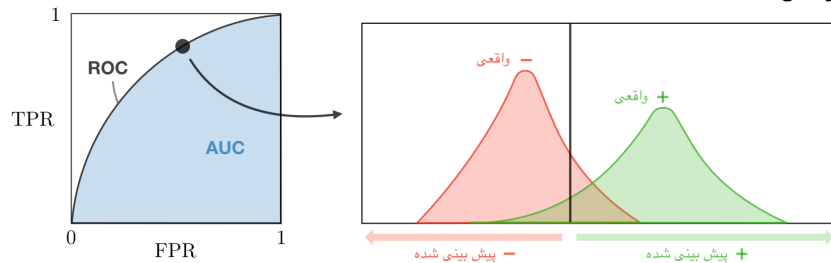


❑ **ROC** – منحنی عملیاتی گیرنده که تحت عنوان ROC نیز شناخته می‌شود تصویر TPR به ازای FPR و با تغییر مقادیر آستانه است. این معیارها بصورت خلاصه در جدول زیر آورده شده‌اند:

معیار	فرمول	معادل
True Positive Rate TPR	$\frac{TP}{TP + FN}$	فراخوانی
False Positive Rate FPR	$\frac{FP}{TN + FP}$	ویژگی-۱

❑ **AUC** – ناحیه‌ی زیر منحنی عملیاتی گیرنده، که با AUC یا AUROC نیز شناخته می‌شود، مساحت زیر منحنی ROC که در شکل زیر نشان داده شده است:



معیارهای وایزش

❑ **معیارهای ابتدایی** – با توجه به مدل وایزش f ، معیارهای زیر برای ارزیابی عملکرد مدل مورد استفاده قرار می‌گیرند:

مجموع کل مربعات	مجموع مربعات توضیح داده شده	باقی‌مانده‌ی مجموع مربعات
$SS_{tot} = \sum_{i=1}^m (y_i - \bar{y})^2$	$SS_{reg} = \sum_{i=1}^m (f(x_i) - \bar{y})^2$	$SS_{res} = \sum_{i=1}^m (y_i - f(x_i))^2$

❑ **ضریب تعیین** – ضریب تعیین، که با R^2 یا r^2 هم نمایش داده می‌شود، معیاری برای سنجش این است که مدل به چه اندازه می‌تواند نتایج مشاهده‌شده را تکرار کند، و به صورت زیر تعریف می‌شود:

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

❑ **معیارهای اصلی** – از معیارهای زیر معمولاً برای ارزیابی عملکرد مدل‌های وایزش با در نظر گرفتن تعداد متغیرهای n که در نظر می‌گیرند، استفاده می‌شود:

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

که L درست‌نمایی و $\hat{\sigma}^2$ تخمینی از واریانس مربوط به هر یک از پاسخ‌ها است.

راهنمای کوتاه نکات و ترفندهای یادگیری ماشین

اقتین عیدی و شروین عیدی

۱۵ شهریور ۱۳۹۸

ترجمه به فارسی توسط الیستر و محمد رضا. بازبینی توسط عرفان نوری و محمد کریمی.

معیارهای دسته‌بندی

❑ **ماتریس درهم‌ریختگی (confusion matrix)** – از ماتریس درهم‌ریختگی برای دست یافتن به تصویری جامع‌تر در ارزیابی عملکرد مدل استفاده می‌شود. این ماتریس بصورت زیر تعریف می‌شود:

دسته پیش‌بینی‌شده

	-	+	
دسته واقعی	FN False Negatives Type II error	TP True Positives	+
	TN True Negatives	FP False Positives Type I error	-

❑ **معیارهای اصلی** – معیارهای زیر معمولاً برای ارزیابی عملکرد مدل‌های دسته‌بندی بکار برده می‌شوند.

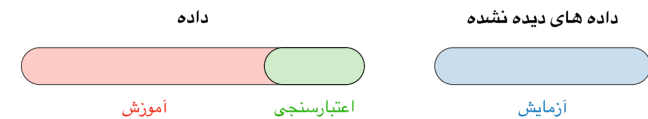
معیار	فرمول	تفسیر
صحت (Accuracy)	$\frac{TP + TN}{TP + TN + FP + FN}$	عملکرد کلی مدل
دقت (Precision)	$\frac{TP}{TP + FP}$	پیش‌بینی‌های مثبت چقدر دقیق هستند
فراخوانی (Recall)	$\frac{TP}{TP + FN}$	پوشش نمونه‌ی مثبت واقعی
ویژگی (Specificity)	$\frac{TN}{TN + FP}$	پوشش نمونه‌ی منفی واقعی
F1 score	$\frac{2TP}{2TP + FP + FN}$	معیار ترکیبی مفید برای دسته‌های نامتوازن

انتخاب مدل

□ **واژگان** – هنگام انتخاب مدل، سه بخش مختلف از داده‌ها را به صورت زیر مشخص می‌کنیم :

مجموعه آموزش (Training)	مجموعه اعتبارسنجی (Validation)	مجموعه آزمایش (Testing)
– مدل آموزش داده شده است – معمولاً ۸۰ درصد از مجموعه داده‌ها	– مدل ارزیابی شده است – معمولاً ۲۰ درصد از مجموعه داده‌ها – این مجموعه همچنین تحت عنوان مجموعه بیرون نگه‌داشته شده یا توسعه نیز شناخته می‌شود	– مدل پیش‌بینی می‌کند – داده‌های دیده نشده

بعد از اینکه مدل انتخاب شد، روی کل مجموعه داده‌ها آموزش داده می‌شود و بر روی مجموعه دادگان دیده نشده آزمایش می‌شود. این مراحل در شکل زیر آمده‌اند :



□ **اعتبارسنجی متقاطع (cross-validation)** – اعتبارسنجی متقاطع، که CV نیز نامیده می‌شود، عبارت است از روشی برای انتخاب مدلی که بیش از حد به مجموعه‌ی آموزش اولیه تکیه نمی‌کند. انواع مختلف بصورت خلاصه در جدول زیر ارائه شده‌اند :

عیب‌شناسی

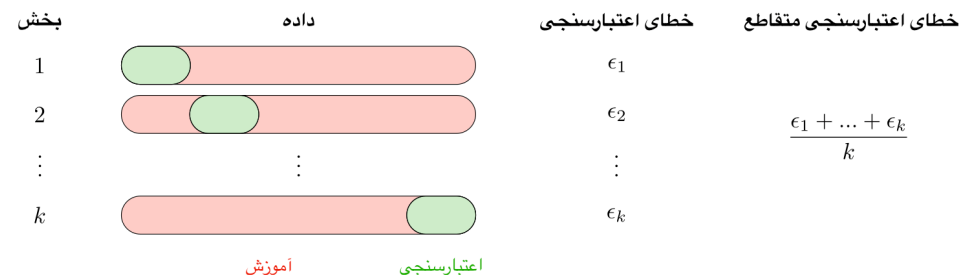
□ **پیش‌قدر (bias)** – پیش‌قدر مدل اختلاف بین پیش‌بینی مورد انتظار و مدل صحیح است که تلاش می‌کنیم برای نمونه داده‌های داده‌شده پیش‌بینی کنیم.

□ **واریانس (variance)** – واریانس یک مدل تنوع پیش‌بینی مدل برای نمونه داده‌های داده‌شده است.

□ **تعادل پیش‌قدر/واریانس** – هر چقدر مدل ساده‌تر باشد، پیش‌قدر بیشتر خواهد بود، و هر چه مدل پیچیده‌تر باشد واریانس بیشتر خواهد شد.

Leave-p-out	k-fold
– آموزش بر روی $n - p$ مشاهده و ارزیابی بر روی p مشاهده‌ی باقی‌مانده – مورد $p = 1$ تحت عنوان حذف تک‌مورد گفته می‌شود	– آموزش بر روی $k - 1$ بخش دیگر و ارزیابی بر روی بخش باقی‌مانده – معمولاً $k = 5$ یا 10

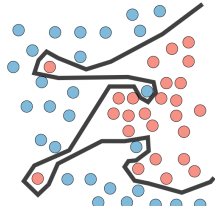
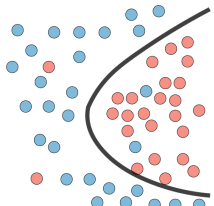
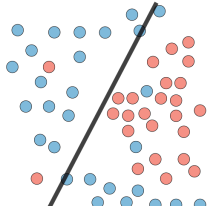
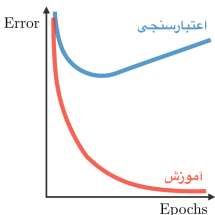
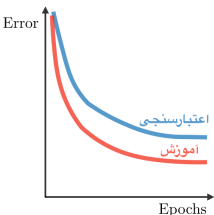
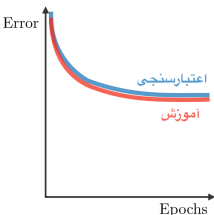
رایج‌ترین روش مورد استفاده، اعتبارسنجی متقاطع k – بخشی نامیده می‌شود که داده‌های آموزشی را به k بخش تقسیم می‌کند تا مدل روی یک بخش ارزیابی شود و در عین حال مدل را روی $k - 1$ بخش دیگر آموزش دهد، و این عمل را k بار تکرار می‌کند. سپس میانگین خطا بر روی k بخش محاسبه می‌شود که خطای اعتبارسنجی متقاطع نامیده می‌شود.



□ **نظام‌بخشی (regularization)** – هدف از رویه‌ی نظام‌بخشی جلوگیری از بیش‌برازش به داده‌ها توسط مدل است و در نتیجه با مشکل واریانس بالا طرف است. جدول زیر خلاصه‌ای از انواع روش‌های متداول نظام‌بخشی را ارائه می‌دهد :

Elastic Net	Ridge	LASSO
بین انتخاب متغیر و ضرایب کوچک مصالحه می‌کند	ضرایب را کوچکتر می‌کند	– ضرایب را تا ۰ کاهش می‌دهد – برای انتخاب متغیر مناسب است
$\dots + \lambda \left[(1 - \alpha) \ \theta\ _1 + \alpha \ \theta\ _2^2 \right]$ $\lambda \in \mathbb{R}, \quad \alpha \in [0, 1]$	$\dots + \lambda \ \theta\ _2^2$ $\lambda \in \mathbb{R}$	$\dots + \lambda \ \theta\ _1$ $\lambda \in \mathbb{R}$

Overfitting	Just right	Underfitting	علامت
– خطای آموزش بسیار کم – خطای آموزش بسیار کم‌تر از خطای آزمایش – واریانس بالا	– خطای آموزش کمی کمتر از خطای آزمایش	– خطای بالای آموزش – خطای آموزش نزدیک به خطای آزمایش – پیش‌قدر زیاد	
			نمایش وایبازش

			نمایش دسته‌بندی
			نمایش یادگیری عمیق
– نظام‌بخشی انجام دهید – داده‌های بیشتری گردآوری کنید		– مدل را پیچیده‌تر کنید – ویژگی‌های بیشتری اضافه کنید – مدت طولانی‌تری آموزش دهید	اصلاحات احتمالی

□ **تحلیل خطا (error analysis)** – تحلیل خطا به بررسی علت اصلی اختلاف در عملکرد بین مدل‌های کنونی و مدل‌های صحیح می‌پردازد.

□ **تحلیل تقطیعی (ablative analysis)** – تحلیل تقطیعی به بررسی علت اصلی اختلاف بین مدل‌های کنونی و مدل‌های پایه می‌پردازد.