

(<https://stanford.edu/%7Eshervine/teaching/cs-229.html>)

## راهنمای کوتاه نکات و ترفندهای یادگیری ماشین

### معیارهای دسته‌بندی

معیارهای اساسی و مهم برای پیگیری در زمینه‌ی دسته‌بندی دوتایی و به منظور ارزیابی عملکرد مدل در زیر آمده‌اند.

ماتریس درهم‌ریختگی - 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	+	<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

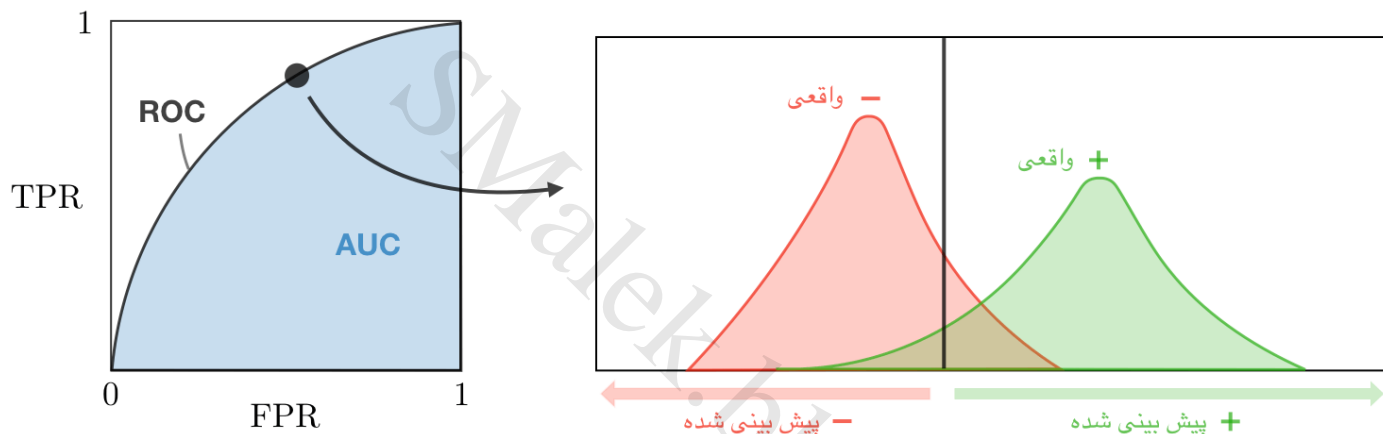
معیارهای اصلی - معیارهای زیر معمولاً برای ارزیابی عملکرد مدل‌های دسته‌بندی بکار برده می‌شوند.

معيار	فرمول
صحت	$\frac{TP + TN}{TP + TN + FP + FN}$
دقت	$\frac{TP}{TP + FP}$
فراخوانی	$\frac{TP}{TP + FN}$
ویژگی	$\frac{TN}{TN + FP}$
F1 score	$\frac{2TP}{2TP + FP + FN}$

**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:

معيار	فرمول	معادل
True Positive Rate TPR	$\frac{TP}{TP + FN}$	فراخوانی
False Positive Rate FPR	$\frac{FP}{TN + FP}$	ویژگی-1

**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
$SS_{\text{tot}} = \sum_{i=1}^m (y_i - \bar{y})^2$	$SS_{\text{reg}} = \sum_{i=1}^m (f(x_i) - \bar{y})^2$	$SS_{\text{res}} = \sum_{i=1}^m (y_i - f(x_i))^2$

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:

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

معیارهای اصلی - از معیارهای زیر معمولاً برای ارزیابی عملکرد مدل‌های واپاش با در نظر گرفتن تعداد متغیرهای  $n$  که در نظر می‌گیرند، استفاده می‌شود:

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

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

## انتخاب مدل

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

مجموعه آموزش	مجموعه اعتبارسنجی	مجموعه آزمایش
--------------	-------------------	---------------

<ul style="list-style-type: none"> <li>• مدل پیش‌بینی می‌کند</li> <li>• داده‌های دیده نشده</li> </ul>	<ul style="list-style-type: none"> <li>• معمولاً ۲۰ درصد از مجموعه داده‌ها</li> <li>• این مجموعه همچنین تحت عنوان مجموعه بیرون نگه‌داشته‌شده یا توسعه نیز شناخته می‌شود</li> </ul>	<ul style="list-style-type: none"> <li>• مدل آموزش داده شده است</li> <li>• معمولاً ۸۰ درصد از مجموعه داده‌ها</li> </ul>
---	--	---




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:



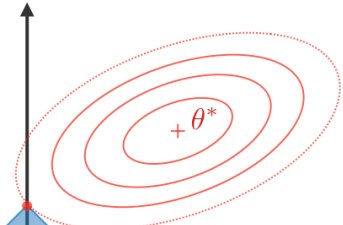
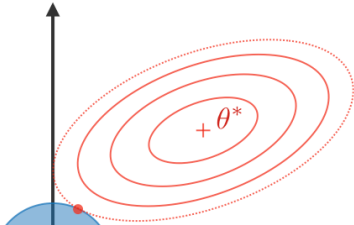
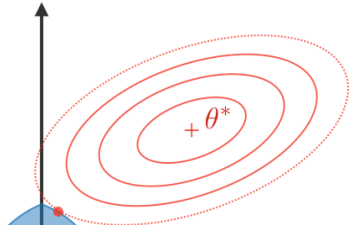
اعتبارسنجی متقاطع – اعتبارسنجی متقاطع، که CV نیز نامیده می‌شود، عبارت است از روشی برای انتخاب مدلی که بیش از حد به مجموعه‌ی آموزش اولیه تکیه نمی‌کند. انواع مختلف بصورت خلاصه در جدول زیر ارائه شده‌اند:

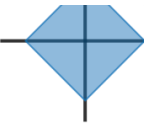
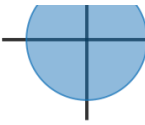
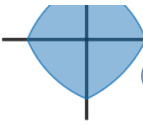
$k$ -fold	Leave- $p$ -out
<ul style="list-style-type: none"> <li>• آموزش بر روی <math>k - 1</math> بخش دیگر و ارزیابی بر روی بخش باقی‌مانده</li> <li>• معمولاً <math>k = 5</math> یا <math>k = 10</math></li> </ul>	<ul style="list-style-type: none"> <li>• آموزش بر روی <math>n - p</math> مشاهده و ارزیابی بر روی <math>p</math> مشاهده‌ی باقی‌مانده</li> <li>• مورد <math>p = 1</math> تحت عنوان حذف تک‌مورد گفته می‌شود</li> </ul>

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.

بخش	داده	خطای اعتبارسنجی	خطای اعتبارسنجی متقاطع
1		$\epsilon_1$	$\frac{\epsilon_1 + \dots + \epsilon_k}{k}$
2		$\epsilon_2$	
$\vdots$	$\vdots$	$\vdots$	
$k$		$\epsilon_k$	
	آموزش	اعتبارسنجی	

نظام‌بخشی - 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:

LASSO	Ridge	Elastic Net
<ul style="list-style-type: none"> <li>• ضرایب را تا ۰ کاهش می‌دهد</li> <li>• برای انتخاب متغیر مناسب است</li> </ul>	<ul style="list-style-type: none"> <li>• ضرایب را کوچکتر می‌کند</li> </ul>	<ul style="list-style-type: none"> <li>• بین انتخاب متغیر و ضرایب کوچک مصالحه می‌کند</li> </ul>
		

 $  \theta  _1 \leq 1$	 $  \theta  _2 \leq 1$	 $(1 - \alpha)  \theta  _1 + \alpha  \theta  _2^2 \leq 1$
$\dots + \lambda   \theta  _1$ $\lambda \in \mathbb{R}$	$\dots + \lambda   \theta  _2^2$ $\lambda \in \mathbb{R}$	$\dots + \lambda \left[ (1 - \alpha)  \theta  _1 + \alpha  \theta  _2^2 \right]$ $\lambda \in \mathbb{R}, \alpha \in [0, 1]$

## عیب‌شناسی

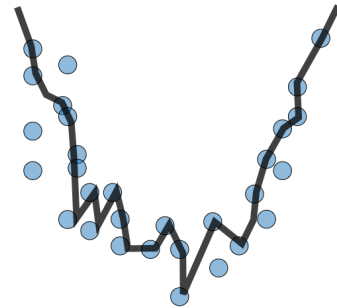
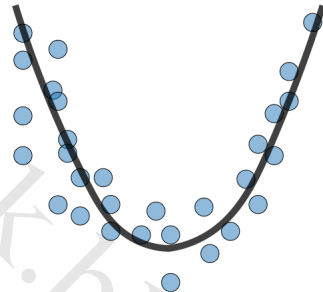
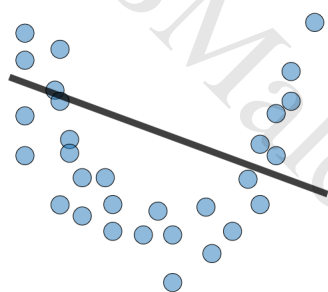
پیش‌قدر - پیش‌قدر مدل اختلاف بین پیش‌بینی مورد انتظار و مدل صحیح است که تلاش می‌کنیم برای نمونه داده‌های داده‌شده پیش‌بینی کنیم.

واریانس - واریانس یک مدل تنوع پیش‌بینی مدل برای نمونه داده‌های داده‌شده است.

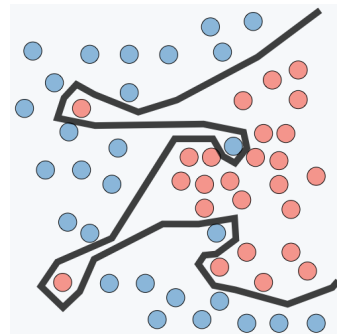
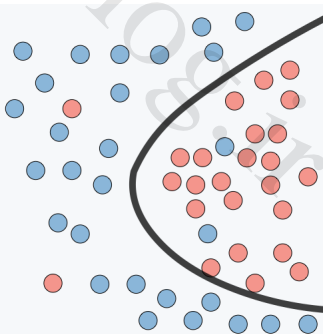
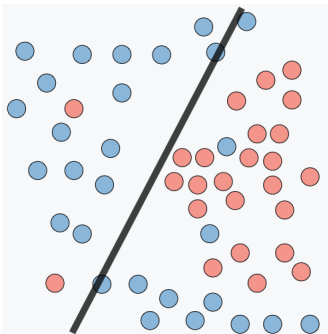
تعادل پیش‌قدر/واریانس - هر چقدر مدل ساده‌تر باشد، پیش‌قدر بیشتر خواهد بود، و هر چه مدل پیچیده‌تر باشد واریانس بیشتر خواهد شد.

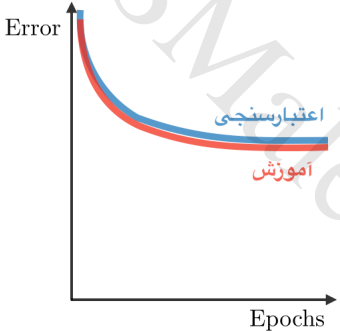
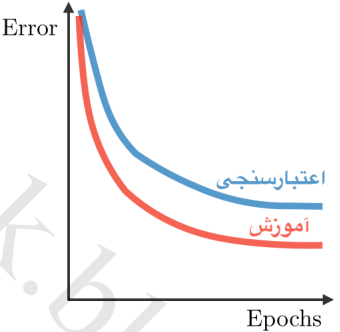
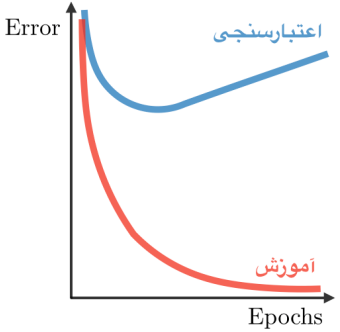
	Underfitting	Just right	Overfitting
علائم	<ul style="list-style-type: none"> <li>خطای بالای آموزش</li> <li>خطای آموزش نزدیک به خطای آزمایش</li> </ul>	<ul style="list-style-type: none"> <li>خطای آموزش کمی کمتر از خطای آزمایش</li> </ul>	<ul style="list-style-type: none"> <li>خطای آموزش بسیار کم</li> <li>خطای آموزش بسیار کمتر از خطای آزمایش</li> </ul>

نمایش وایازش



نمایش دسته‌بندی



نمایش یادگیری عمیق	 <p>The graph shows 'Error' on the y-axis and 'Epochs' on the x-axis. A red line labeled 'آموزش' (Training) and a blue line labeled 'اعتبارسنجی' (Validation) both decrease and then plateau at a low error level.</p>	 <p>The graph shows 'Error' on the y-axis and 'Epochs' on the x-axis. The red line labeled 'آموزش' (Training) decreases, while the blue line labeled 'اعتبارسنجی' (Validation) increases after an initial decrease.</p>	 <p>The graph shows 'Error' on the y-axis and 'Epochs' on the x-axis. The red line labeled 'آموزش' (Training) decreases sharply, while the blue line labeled 'اعتبارسنجی' (Validation) increases significantly after a small initial decrease.</p>
اصلاحات احتمالی	<ul style="list-style-type: none"> <li>• مدل را پیچیده‌تر کنید</li> <li>• ویژگی‌های بیشتری اضافه کنید</li> <li>• مدت طولانی‌تری آموزش دهید</li> </ul>		<ul style="list-style-type: none"> <li>• نظام‌بخشی انجام دهید</li> <li>• داده‌های بیشتری گردآوری کنید</li> </ul>

تحلیل خطا - تحلیل خطا به بررسی علت اصلی اختلاف در عملکرد بین مدل‌های کنونی و مدل‌های صحیح می‌پردازد.

تحلیل تقطیعی - تحلیل تقطیعی به بررسی علت اصلی اختلاف بین مدل‌های کنونی و مدل‌های پایه می‌پردازد.