



موضوع

معیار های سنجش در یک مدل طبقه بندی

ارائه دهنده

سارا معصومی

آبان ۱۴۰۱

ماتریس در هم ریختگی (Confusion Matrix)

جدول و یا ماتریس درهم ریختگی، نتایج حاصل از طبقه‌بندی را بر اساس اطلاعات واقعی موجود، نمایش می‌دهد. سطرهای این ماتریس کلاس‌های موجود در اطلاعات واقعی (مشاهدات) است و ستون‌های این ماتریس کلاس‌های موجود در زمان پیش‌بینی است. همانطور که میدانیم تعداد کلاس‌ها چه در اطلاعات واقعی و چه در زمان پیش‌بینی یکسان است بنابراین این ماتریس یک ماتریس مربعی می‌باشد. قطر اصلی این ماتریس تعداد پیش‌بینی‌های منطبق بر واقعیت هستند و خارج قطر اصلی تعداد پیش‌بینی‌هایی که با واقعیت منطبق نیستند و در واقع خطای پیش‌بینی می‌باشند. چنانچه صحت مدل ۱۰۰٪ باشد به این معنا که تمام پیش‌بینی‌های صورت گرفته منطبق بر واقعیت باشند آنگاه ماتریس در هم ریختگی ما یک ماتریس قطری خواهد بود که البته در عمل چنین صحتی برای مدل وجود ندارد.

		Predicted Values	
		Positive	Negative
Actual Values	Positive	TP	FN
	Negative	FP	TN

مفاهیم

- مقادیر مثبت صحیح (TP): نمونه‌ی مربوطه در مشاهدات عضو دسته‌ی مثبت باشد و طبق پیش‌بینی نیز عضو دسته‌ی مثبت ثبت شود.
- مقادیر منفی کاذب (FN): نمونه‌ی مربوطه در مشاهدات عضو دسته‌ی مثبت باشد ولی طبق پیش‌بینی عضو دسته‌ی منفی ثبت شود.
- مقادیر مثبت کاذب (FP): نمونه‌ی مربوطه در مشاهدات عضو دسته‌ی منفی باشد ولی طبق پیش‌بینی عضو دسته‌ی مثبت ثبت شود.
- مقادیر منفی صحیح (TN): نمونه‌ی مربوطه در مشاهدات عضو دسته‌ی منفی باشد و طبق پیش‌بینی نیز عضو دسته‌ی منفی ثبت شود.

مثال

فرض کنید از افراد مراجعه کننده به یک مرکز ترک اعتیاد طی یک روز اسکن ریه دریافت شده است و همینطور سوالی در مورد مصرف سیگار آنها مطرح شده است که جواب ها در غالب پاسخ های "بله" یا "خیر" ثبت شده اند.

مقادیر پیش بینی افراد سیگاری

		بله	خیر
مقادیر واقعی افراد سیگاری	بله	20	2
	خیر	9	69

معیارهای سنجش اعتبار مدل در طبقه بندی

پس از تعریف ماتریس درهم ریختگی و مفاهیم ذکر شده حال میتوان با استفاده از آنها معیار هایی را جهت سنجش اعتبار نتایج حاصل از مدل معرفی نمود. این معیار ها عبارت اند از :

- **صحت (Accuracy):** صحت متداول ترین معیار سنجش میباشد. درواقع صحت میزان صحیح بودن نتایج حاصل از پیش بینی مدل را طبق واقعیت برای تمام کلاس ها اندازه گیری میکند.

$$\text{Accuracy} = (TP+TN) / (TP+FN+FP+TN)$$

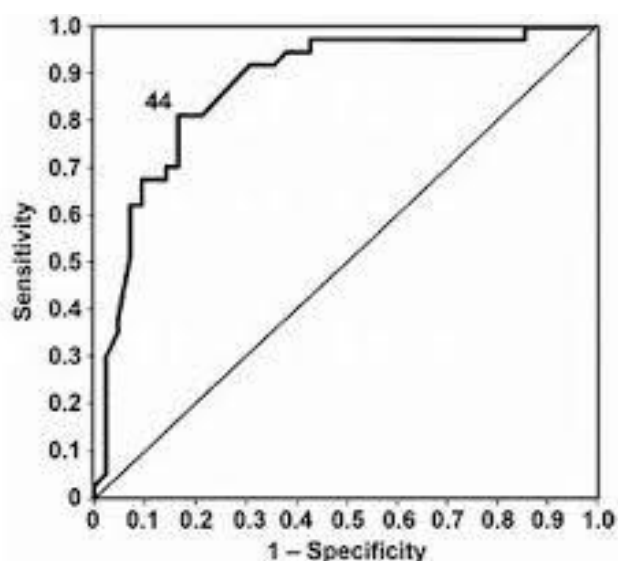
- **حساسیت (Sensitivity):** زمانی که پژوهشگر از این پارامتر به عنوان پارامتر ارزیابی برای دسته بندی استفاده می کند، هدفش دستیابی به نهایت صحت در تشخیص فقط نمونه های کلاس مثبت است. حساسیت را "نرخ پاسخ های مثبت درست" نیز می گویند. حساسیت به معنی نسبتی از موارد مثبت است که مدل آن ها را به درستی به عنوان نمونه مثبت تشخیص داده است.

$$\text{Sensitivity (TPR)} = TP / (TP+FN)$$

- **خاصیت / ویژگی (Specificity):** در نقطه مقابل حساسیت، ممکن است در مواقعی صحت تشخیص کلاس منفی حائز اهمیت باشد. از متداول ترین پارامترها که معمولاً در کنار حساسیت بررسی می‌شود، پارامتر خاصیت، است که به آن "نرخ پاسخ‌های منفی درست" نیز می‌گویند. خاصیت به معنی نسبتی از موارد منفی است که مدل آن‌ها را به درستی به عنوان نمونه منفی تشخیص داده است.

$$\text{Specificity (TNR)} = \text{TN} / (\text{TN} + \text{FP})$$

- **منحنی مشخصه عملکرد سیستم ROC:** منحنی که ارتباط بین دو پارامتر حساسیت و خاصیت را بیان می‌کند. محور عمودی این نمودار نشان‌دهنده نرخ مثبت صحیح (Sensitivity)، و محور افقی نشان‌دهنده مقدار نرخ مثبت غلط (1-Specificity) است. نتایج مختلف دسته‌بندی نشانگر نقاط مختلف بر روی این نمودار هستند و در نهایت یک منحنی را تشکیل می‌دهند. با توجه به شکل زیر، در بهترین حالت و با فرض طبقه‌بندی صد درصد صحیح در هر دو دسته، نقطه مربوطه عبارت است از نقطه گوشه بالای سمت چپ، یعنی نقطه (0,1) و نیز با فرض دسته‌بندی به صورت تصادفی، نقطه متناظر در منحنی، یکی از نقاط موجود روی خط واصل نقطه (0,0) و نقطه (1,1) خواهد بود. در واقعیت، منحنی حاصل از یک دسته‌بندی، منحنی بین این دو حالت است. مساحت زیر این نمودار (Area Under Curve)، به عنوان یک معیار برای ارزیابی عملکرد دسته‌بند مورد استفاده قرار می‌گیرد. با توجه به توضیحاتی که پیش‌تر ارائه شد، بدیهی است که در حالت ایده‌آل، مساحت زیر منحنی برابر با بیشترین مقدار خود، یعنی یک است. بنابراین، هر چه مساحت زیر نمودار به عدد یک نزدیک‌تر باشد، به معنای بهتر بودن عملکرد دسته‌بند است.



- **دقت (Precision):** نسبت پاسخ های مثبت صحیح به تمام پاسخ های مثبت پیش بینی شده میباشد.

$$PPV = TP / (TP + FP)$$

- **فراخوانی / یادآوری (Recall):** در مقابل دقت معیار فراخوانی نسبت پاسخ های منفی صحیح به کل پاسخ های منفی پیش بینی شده میباشد.

$$NPV = TN / (TN + FN)$$

- **معیار F1:** اگر بخواهیم همزمان هر دو معیار صحت و فراخوانی در ارزیابی مدل دخیل باشند از این معیار استفاده می کنیم. برای مثال زمانی که بخواهیم بدانیم پیش بینی مدل ما تا چه اندازه با واقعیت مطابقت دارد و تا چه اندازه به درستی پیش بینی را برای هر دو کلاس انجام داده است.

$$F\text{-measure} = 2 * (Recall * Precision) / (Recall + Precision)$$

نکات

- معیار های صحت ، حساسیت و خاصیت به صورت درصد بیان میشوند.
- واضح است که پیش بینی عالی، پیش بینی است که مقادیر Sensitivity و Specificity مربوط به آن، هر دو صد درصد باشند؛ اما احتمال وقوع این اتفاق در واقعیت بسیار کم است و همیشه یک حداقل خطایی وجود دارد.

تعمیم معیار های ذکر شده به مسائلی با بیش از دو کلاس

مثال: فرض کنید یک مدل طبقه بندی با ۴ کلاس مجزا برازش داده شده است. و ماتریس در هم ریختگی آن به صورت زیر میباشد.

	A	B	C	D
A	12	0	0	1
B	1	8	1	0
C	0	0	9	1
D	0	2	0	10

ردیف اول:

$$TP = 12 \quad TN = 8+9+10 \quad FN = 1 \quad FP = 1$$

ردیف دوم:

$$TP = 8 \quad TN = 12+9+10 \quad FN = 1+1 \quad FP = 2$$

ردیف سوم:

$$TP = 9 \quad TN = 12+8+10 \quad FN = 1 \quad FP = 1$$

ردیف اول:

$$TP = 10 \quad TN = 12+8+9 \quad FN = 2 \quad FP = 2$$

$$\sum TP = 39 \quad \sum TN = 117 \quad \sum FN = 6 \quad \sum FP = 6$$

$$ACC = \frac{TP+TN}{TP+TN+FN+FP} = \frac{39 + 117}{39 + 117 + 6 + 6} = 0.92$$

❖ مقاله های بیشتر در <https://github.com/Sara0M>