

# یادگیری ماشین برای بیوانفورماتیک

محمدرضا شمس اشکذری

۱ اسفند ۱۳۹۷

۹۵۱۰۰۱۵۶

۱ ۱

با توجه به تعریف هر کدام از موارد مطرح شده، در جدول زیر حاصل محاسبات را برای هر کدام از الگوریتم ها آورده ایم:

۲ ۲

از نظر شهودی بیشتر به نظرم درستی الگوریتم مورد نظر ماست و اینکه اگر فردی مبتلا هست، به اشتباه جواب ندهیم. لذا می توان گفت بعد از درستی به سراغ این می رویم که تا چه حد ما مثبت ها را درست پیش بینی کرده ایم که این دقیقا تعریف همان precision است.

۳ ۳

به طور کلی به نظرم درستی در جایی مفید واقع می شود که داده ها جنبه پیوسته نداشته باشند یا حداقل متر ما برای اندازه گیری خطا پیوسته نباشد. زیرا فرضا در کاری مثل تشخیص چهره، چهره واقعی فقط یکی است و در این صورت درستی نسبت ۱ به کل داده ها خواهد بود که اصلا مفید

	accuracy	recall	precision
alg1	0.967	0.8	0.47
alg2	0.927	0.93	0.28

نیست در این حالت. حال اگر همان تشخیص چهره را با یک محدوده خطا در نظر بگیریم و گسسته سازی کنیم، درستی چیز بهتری نشان خواهد داد یا اینکه مثلاً مساله را تشخیص سفید پوست از سیاه پوست بگیریم.

زمانی که ارزش positives false بالا باشد، معیار صحت، معیار مناسبی خواهد بود. فرض کنید، مدلی برای تشخیص سرطان داشته باشیم و این مدل Precision پایینی داشته باشد. نتیجه این امر این است که این مدل، بیماری بسیاری از افراد را به اشتباه سرطان تشخیص می‌دهد. نتیجه این امر استرس زیاد، آزمایش‌های فراوان و هزینه‌های گزافی را برای بیمار به دنبال خواهد داشت. زمانی که ارزش negatives false بالا باشد، معیار Recall، معیار مناسبی خواهد بود. فرض کنیم مدلی برای تشخیص بیماری کشنده ابولا داشته باشیم. اگر این مدل Recall پایینی داشته باشد چه اتفاقی خواهد افتاد؟ این مدل افراد زیادی که آلوده به این بیماری کشنده هستند را سالم در نظر می‌گیرد و این فاجعه است.