

رسالة محمد



یادگیری عمیق

مدرس: محمدرضا محمدی

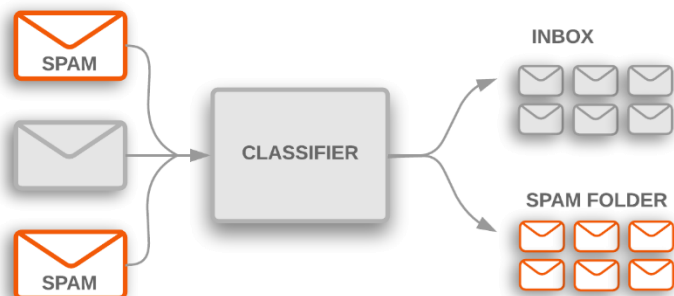
بهار ۱۴۰۲

روش‌شناسی کاربردی

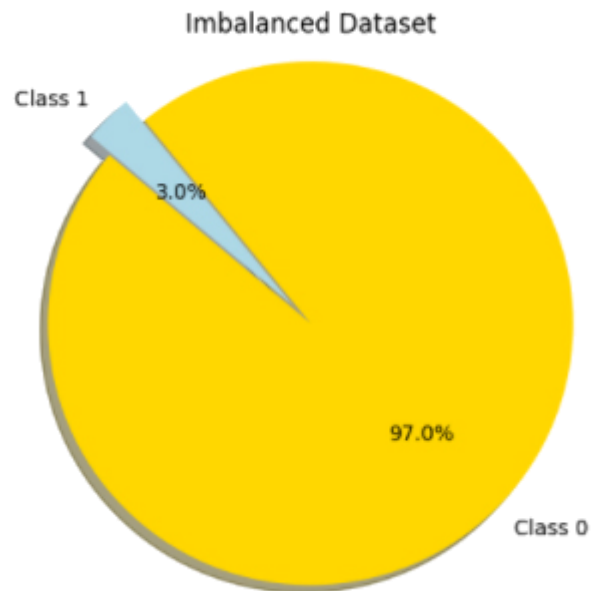
Practical Methodology

معیارهای ارزیابی عملکرد

- از کدام معیار استفاده کنیم؟
- معیارهای ارزیابی مختلفی می‌توانند برای اندازه‌گیری اثربخشی یک برنامه استفاده شوند
- این معیارها معمولاً با تابع ضرر مورد استفاده برای آموزش مدل متفاوت هستند
- گاهی اوقات مرتکب شدن یک نوع خطا از نوع دیگر پرهزینه‌تر است
 - سیستم تشخیص ایمیل هرزنامه
 - دسته‌بندی نادرست یک پیام قانونی به عنوان هرزنامه می‌تواند هزینه بسیار بیشتری داشته باشد



معیارهای ارزیابی عملکرد



- آموزش یک دسته‌بند دوسطحی برای تشخیص برخی رویدادهای نادر

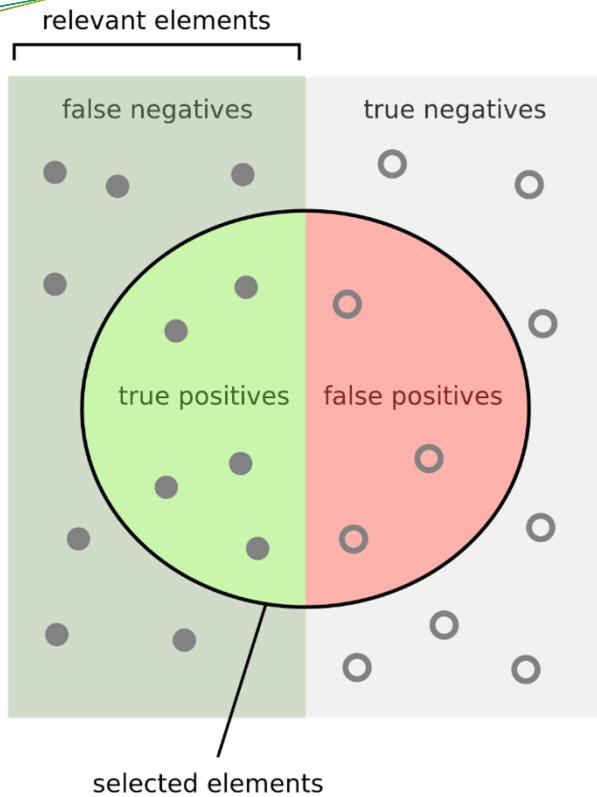
- آزمایش پزشکی برای یک بیماری نادر

- ۰.۱٪ احتمال ابتلا به بیماری

- دقت ۹۹.۹٪ با کدنویسی ساده!

- واضح است که دقت (Accuracy) معیار مناسبی برای توصیف عملکرد چنین مسائلی نیست

Recall و Precision



• Precision:

- درصد نمونه‌هایی که توسط مدل به عنوان کلاس مثبت تشخیص داده شده‌اند و درست بوده‌اند


$$Precision = \frac{TP}{TP + FP}$$

• Recall:


- درصد نمونه‌هایی که مثبت بوده‌اند و به درستی توسط مدل تشخیص داده شده‌اند

$$Recall = \frac{TP}{TP + FN}$$

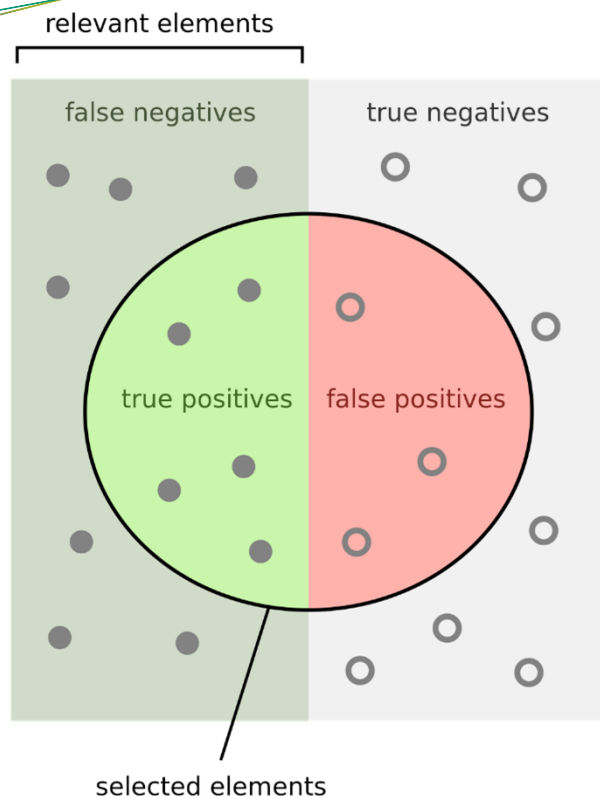
How many selected items are relevant?

Precision = 

How many relevant items are selected?

Recall = 

Recall و Precision



• یک مدل که می گوید هیچ کس این بیماری را ندارد

- Precision: کامل (۱۰۰٪)

- Recall: صفر

• یک مدل که می گوید تمام افراد این بیماری را دارند

- Recall: کامل (۱۰۰٪)

- Precision: برابر با درصد افراد مبتلا به این بیماری (مثلا ۰.۰۰۱)

How many selected items are relevant?

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

Recall و Precision

- خلاصه‌سازی PR با یک عدد

$$F_1 = \frac{2PR}{P + R}$$

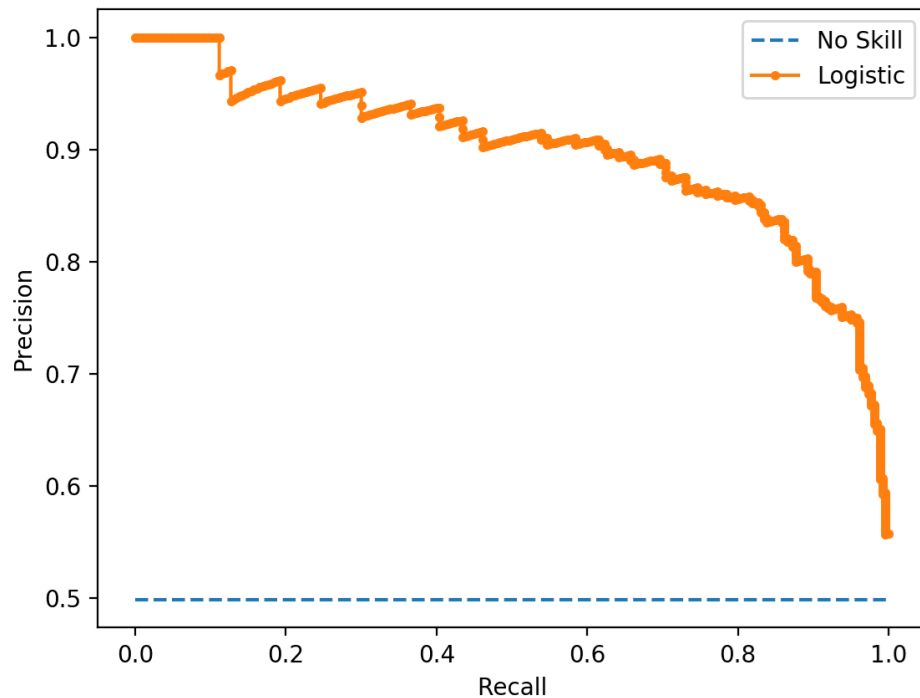
F-score -

- خلاصه‌سازی PR با گراف

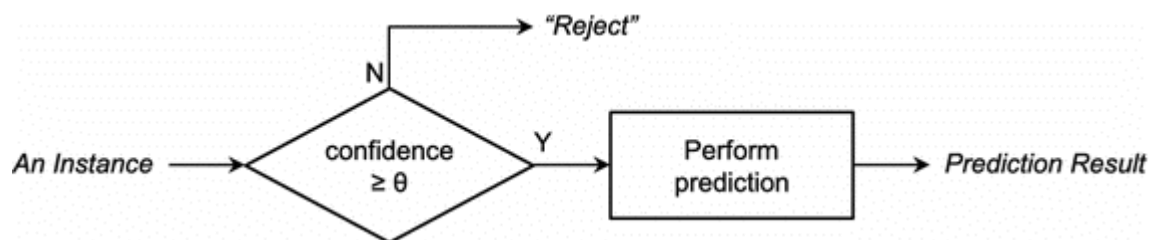
- Precision در محور y و Recall در محور x

- به طور معمول دسته‌بند برای هر کلاس یک امتیاز را محاسبه می‌کند و انتظار می‌رود کلاس درست دارای امتیاز بالاتری باشد

- سطح زیر منحنی (AUC) یکی از معیارهای بسیار پرکاربرد برای خلاصه‌سازی این منحنی است



معیارهای ارزیابی عملکرد



- امتناع از تصمیم‌گیری

- نیاز است تخمین بزنیم که چقدر در مورد تصمیم گرفته شده مطمئن هستیم
- به خصوص اگر یک تصمیم اشتباه می‌تواند ضرر زیاده داشته باشد و اگر یک اپراتور انسانی بتواند گاهی مواقع مسئولیت را بر عهده بگیرد

- کاهش میزان کاری که انسان باید انجام دهد

- پوشش (Coverage)

- محدوده پاسخ سیستم (برای چند درصد تصمیم نمی‌گیرد)

- در پروژه Street View هدف‌گذاری زیر انجام شده است

- دقت ۹۸٪ با پوشش ۹۵٪

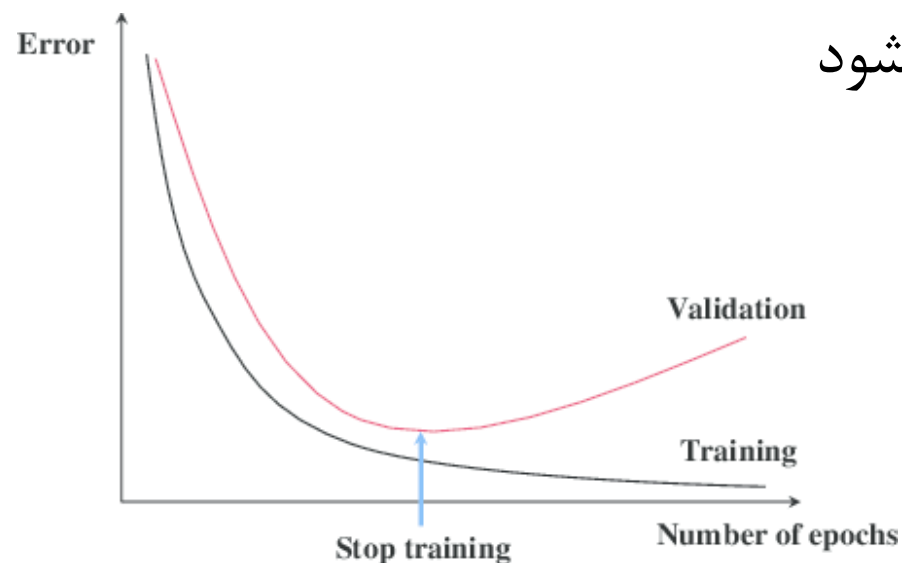


مدل‌های پایه پیش فرض

- یک سیستم end-to-end منطقی پیاده‌سازی کنید
 - یادگیری عمیق به عنوان یک رویکرد end-to-end با کدهای بسیار زیادی که در دسترس است، اجازه می‌دهد تا به راحتی یک الگوریتم پایه را پیاده‌سازی کنیم
- توصیه‌هایی برای انتخاب الگوریتم پایه
 - بسته به پیچیدگی مسئله
 - مدل یادگیری با عمق زیاد یا کم
 - بسته به ساختار داده‌ها
 - استفاده از لایه‌های کاملاً متصل و/یا لایه‌های کانولوشنی و/یا لایه‌های بازگشتی و/یا لایه‌های توجه
- بهینه‌ساز Adam با مقادیر $\text{beta1} = 0.9$ و $\text{beta2} = 0.999$ و $\text{learning_rate} = 1\text{e-}3$ or $5\text{e-}4$ یک نقطه شروع خوب است

مدل‌های پایه پیش فرض

- اگر مجموعه آموزشی محدود است
 - می‌توان از همان ابتدا از روش‌های منظم‌سازی استفاده کرد
 - Dropout، داده‌افزایی، منظم‌سازی وزن‌ها و ...
 - نرمال‌سازی می‌تواند منجر به بهبود تعمیم‌دهی شود
- توقف زودهنگام (Early Stopping) تقریباً به طور کلی باید استفاده شود
- اگر مسئله مورد نظر شبیه به مسئله شناخته شده دیگری است
 - مدل و الگوریتمی که در حال حاضر بهترین عملکرد را دارد کپی کنید
 - از وزن‌های پیش‌آموخته به درستی استفاده کنید

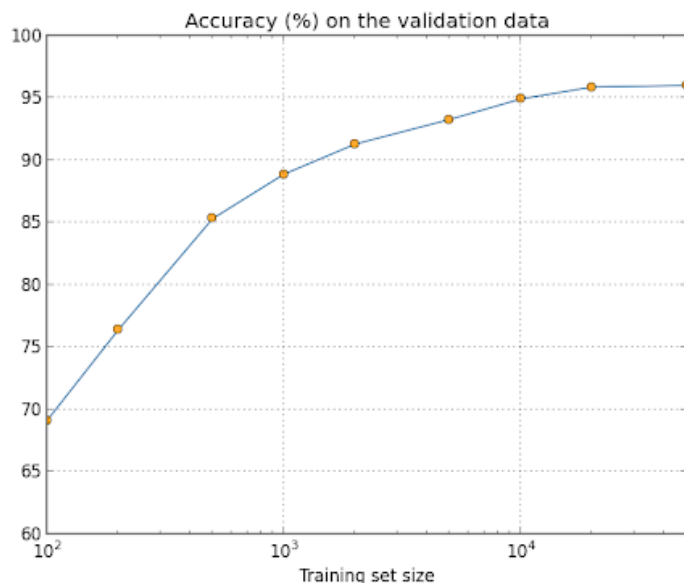


جمع‌آوری داده‌های بیشتر

- چگونه می‌توان تصمیم گرفت که داده‌های بیشتری را جمع‌آوری کرد؟
 - آیا عملکرد روی مجموعه آموزشی قابل قبول است یا خیر؟
- اگر عملکرد در مجموعه آموزشی ضعیف باشد
 - دلیلی برای جمع‌آوری داده‌های بیشتر وجود ندارد
 - ظرفیت مدل را افزایش دهید
 - الگوریتم یادگیری را بهبود دهید
- اگر مدل‌های قوی به خوبی کار نمی‌کنند
 - ممکن است اشکال منطقی در کد وجود داشته باشد
 - ممکن است مشکل از کیفیت داده‌های آموزشی باشد
 - سعی کنید داده‌های تمیزتر جمع‌آوری کنید یا مجموعه کامل‌تری از ویژگی‌ها را جمع‌آوری کنید

جمع‌آوری داده‌های بیشتر

- اگر عملکرد در مجموعه آزمون قابل قبول است
 - نیاز به کار دیگری نیست!
- اگر عملکرد در مجموعه آزمون بسیار بدتر از مجموعه آموزشی
 - جمع‌آوری داده‌های بیشتر یکی از موثرترین راه‌حل‌ها است
- ملاحظات کلیدی



- هزینه و امکان‌سنجی جمع‌آوری داده‌های بیشتر
- هزینه و امکان‌سنجی کاهش خطای آزمون با روش‌های دیگر
- تخمین مقدار داده‌ای که انتظار می‌رود ضروری باشد