



یادگیری ماشین

تمرین «یک»

دستیاران آموزشی

روشنک طالشی

امیرمحمد کوبش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه های هوشمند

نیم سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۲۹ مهر ۱۴۰۲

۱. $[1R - 30\%]$ هدف از این تمرین پیاده سازی الگوریتم OneRule است.

در این الگوریتم، قاعده ای ساده توسط الگوریتم ایجاد می شود که بر اساس آن، هر نمونه داده به یکی از دسته ها تعلق می گیرد. پس خروجی اصلی در این تکلیف استخراج یک قاعده تصمیم گیری است. در ادامه، به توضیح بخش های مختلف این سوال می پردازیم:

▪ پیش پردازش داده

قدم بعدی در حل سوال این است که داده ها را پیش پردازش کنید تا مشکلات موجود در آن ها شناسایی و رفع شوند. این مرحله شامل بررسی و پردازش داده های از دست رفته، مقادیر نامعتبر و همچنین داده های عددی هستند. برای پاکسازی داده های نامعلوم می توانید آن ها را با مقادیر مناسب مانند مینیمم یا میانگین جایگزین کنید و در صورت لزوم آن سطر را حذف کنید. همچنین در صورت لزوم می توانید با روش هایی مقادیر عددی را نیز به اسمی تبدیل کنید. راه حل شما برای کار با ستون هایی با محتوای عددی چیست؟

▪ پیاده سازی الگوریتم OneRule

در این بخش از شما انتظار داریم که یک در پایتون یک کلاس به نام OneRule تعریف کنید که دارای دو تابع اصلی `fit` و `predict` و دیگر توابع است.

تابع `fit`: در این بخش، قاعده ها بر اساس داده های آموزشی تولید می شوند. الگوریتم سعی می کند با تحلیل داده های آموزشی یک قاعده ساده تعیین کند که بر اساس آن، دسته بندی موثری بر روی داده ها انجام شود. در ورودی داده های آموزش وارد میشوند و خروجی قاعده تولید شده است. به ازای هر ستون `accuracy` محاسبه شده را نیز گزارش کنید.

Algorithm 1 Rule Extraction from Training Data

```
1: Input: Training data  $D$ 
2: Output: Rule  $R$ 
3: for  $i \in \{1, \dots, n\}$  do ▷ number of attributes
4:   for  $v \in \{1, \dots, m\}$  do ▷ number of values of attribute  $i$ 
5:      $c_i \leftarrow \text{count}(D_i)$  ▷ number of instances of class  $c$  in  $D_i$ 
6:      $c \leftarrow \max(c_i)$  ▷ most frequent class in  $D_i$ 
7:      $R \leftarrow \text{assign}(R, c, i, v)$  ▷ assign class  $c$  to attribute  $i$  with value  $v$ 
8:   end for
9: end for
10:  $e \leftarrow \text{error}(R, D)$  ▷ error rate of rule  $R$  on data  $D$ 
11: Return  $R$  ▷ rule with smallest error rate
```

شکل ۱: الگوریتم 1R



یادگیری ماشین

تمرین «یک»

دستیاران آموزشی

روشنک طالشی

امیرمحمد کوشش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه های هوشمند

نیم سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۲۹ مهر ۱۴۰۲

تابع `predict` در این تابع، داده های تست با استفاده از قاعده های تولید شده در تابع `fit`، دسته بندی می شوند. این تابع وظیفه پیش بینی دسته بندی داده های تست را بر عهده دارد. در ورودی، داده تست و قاعده را بدهید و در خروجی امتیازهای ذکر شده اعلام شود.

▪ آشنایی با داده

در این بخش، دادگان `HR-Employee-Attrition.csv` را تصادفی به دو بخش تقسیم کنید: ۸۰٪ برای آموزش و ۲۰٪ برای آزمایش. این تقسیم برای ارزیابی عملکرد الگوریتم `One Rule` استفاده می شود. داده های آموزش برای ساخت قاعده و داده های آزمایش برای ارزیابی عملکرد و دقت الگوریتم استفاده می شوند.

در پایان سوال توضیحاتی در مورد دادگان ارائه شده آورده شده است.

▪ ارزیابی

در ادامه ماتریس درهم ریختگی^۱ را رسم کنید (در ادامه توضیحاتی در این باره آورده ایم) و نهایتاً با استفاده از معیارهایی مانند `precision`، `recall`، `F-score` و `ROC` عملکرد الگوریتم `OneRule` بر روی داده های آزمایش بررسی شود. این معیارها به عنوان معیارهای ارزیابی استفاده می شوند تا دقت و صحت دسته بندی الگوریتم را نشان دهند. نتایج این ارزیابی، میزان عملکرد و قابلیت پیش بینی الگوریتم `OneRule` را نشان می دهد. مزایا و محدودیت های این الگوریتم چیست؟ آیا راهی برای بهبود محدودیت های آن سراغ دارید؟ در گزارش شرح دهید.

ماتریس درهم ریختگی یک ابزار اندازه گیری عملکرد است که در یادگیری ماشین و آمار برای ارزیابی دقت یک مدل طبقه بندی استفاده می شود. این جدولی است که پیش بینی های انجام شده توسط مدل را در برابر برچسب های کلاس واقعی داده ها خلاصه می کند.

جدول ۱: ساختار ماتریس درهم ریختگی

		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

^۱ Confusion Matrix



یادگیری ماشین

تمرین «یک»

دستیاران آموزشی

روشنک طالشی

امیرمحمد کوبش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه های هوشمند

نیم سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۲۹ مهر ۱۴۰۲

در ماتریس درهم ریختگی، ردیف ها برچسب های کلاس واقعی را نشان می دهند، در حالی که ستون ها نشان دهنده برچسب های کلاس پیش بینی شده هستند. چهار عبارت موجود در ماتریس دارای تفسیر زیر هستند:

- مثبت واقعی^۱: مدل کلاس مثبت را به درستی پیش بینی کرد.
- مثبت کاذب^۲: مدل به اشتباه کلاس مثبت را زمانی که کلاس واقعی منفی بود (خطای نوع I) پیش بینی کرد.
- منفی کاذب^۳: مدل به اشتباه کلاس منفی را زمانی که کلاس واقعی مثبت بود (خطای نوع II) پیش بینی کرد.
- منفی واقعی^۴: مدل کلاس منفی را به درستی پیش بینی کرد.

دقت، یادآوری و امتیاز F1 معمولاً معیارهایی هستند که از ماتریس درهم ریختگی برای ارزیابی عملکرد یک مدل طبقه بندی استفاده می شوند. آنها به صورت زیر محاسبه می شوند:

- **Precision**: دقت نسبت مثبت های واقعی به مجموع مثبت های واقعی و مثبت های کاذب است. دقت پیش بینی های مثبت مدل را اندازه گیری می کند. فرمول (۱) نحوه محاسبه این روش را نشان می دهد.

$$Recall = \frac{TP}{TP + FP} \quad (1)$$

دقت بر نسبت پیش بینی های مثبتی که واقعاً درست هستند، تمرکز می کند و بینشی در مورد توانایی مدل برای اجتناب از مثبت های کاذب ارائه می کند.

- **Recall**: Recall که به عنوان حساسیت یا نرخ مثبت واقعی نیز شناخته می شود، نسبت مثبت های واقعی به مجموع مثبت های واقعی و منفی های کاذب است. توانایی مدل در شناسایی صحیح موارد مثبت را می سنجد.

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

¹ True Positive

² True Negative

³ False Negative

⁴ True Negative



یادگیری ماشین

تمرین «یک»

دستیاران آموزشی

روشنگ طالشی

امیرمحمد کوشش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۲۹ مهر ۱۴۰۲

Recall بر نسبت نمونه‌های مثبتی که به درستی شناسایی شده‌اند، تأکید می‌کند، که نشان‌دهنده توانایی مدل برای اجتناب از منفی‌های کاذب است.

- **امتیاز F1:** امتیاز F1 میانگین هارمونیک دقت و یادآوری است. اندازه‌گیری متعادلی را ارائه می‌دهد که هم precision و هم Recall را به طور همزمان در نظر می‌گیرد.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

امتیاز F1، precision و Recall را در یک متریک واحد ترکیب می‌کند، که زمانی مفید است که توزیع کلاس نامتعادل باشد.

- **منحنی مشخصه عملیاتی گیرنده (ROC):** یک نمایش گرافیکی است که عملکرد یک مدل طبقه‌بندی باینری را در آستانه‌های طبقه‌بندی مختلف نشان می‌دهد. نرخ مثبت واقعی^۱ را در برابر نرخ مثبت کاذب^۲ در تنظیمات آستانه‌های مختلف ترسیم می‌کند.
- فرمول FPR و TPR به صورت زیر است:

$$FPR = \frac{FP}{FP + TN} \quad (4)$$

$$TPR = \frac{TP}{TP + FN} \quad (5)$$

منحنی ROC با نشان دادن اینکه چگونه عملکرد مدل با مقادیر آستانه متفاوت تغییر می‌کند، به ارزیابی مبادله بین مثبت واقعی و مثبت کاذب کمک می‌کند. ناحیه زیر منحنی (AUC-ROC) ROC اغلب به عنوان یک متریک خلاصه برای تعیین کمیت عملکرد کلی یک طبقه‌بندی کننده استفاده می‌شود. AUC-ROC بالاتر نشان دهنده توانایی تشخیص بهتر مدل است.

ستون‌های داده‌ی کارکنان منابع انسانی به شرح زیر است

▪ **Age:** این ستون نشان دهنده سن کارکنان است.

¹ True Positive Rate

² False Positive Rate



یادگیری ماشین

تمرین «یک»

دستیاران آموزشی

روشنک طالشی

امیرمحمد کوشش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۲۹ مهر ۱۴۰۲

- **Attrition**: این ستون نشان می‌دهد که آیا کارمندی شرکت را ترک کرده است یا هنوز مشغول به کار است. این ستون حاوی مقادیر باینری مانند «بله» یا «خیر» برای نشان دادن وضعیت فرسایشی است. در این داده این ستون را به عنوان کلاس مدنظر قرار داده‌ایم.
- **BusinessTravel**: این ستون تعداد یا نوع سفرهای کاری انجام شده توسط کارمندان را نشان می‌دهد. شامل دسته‌هایی مانند «غیر مسافرتی»، «به ندرت سفر کنید» یا «مکرر سفر کنید».
- **DailyRate**: این ستون به نرخ یا حقوق روزانه کارکنان اشاره دارد.
- **Department**: این ستون بخش یا ناحیه عملکردی را مشخص می‌کند که کارکنان در آن کار می‌کنند. شامل بخش‌هایی مانند «فروش»، «منابع انسانی» یا «تحقیق و توسعه» باشد.
- **DistanceFromHome**: این ستون فاصله بین خانه کارمند و محل کارشان را بر حسب مایل نشان می‌دهد.
- **Education**: این ستون نشان دهنده بالاترین سطح تحصیلات است که توسط کارکنان کسب شده است. شامل مقولاتی مانند "دبیرستان"، "مدرک لیسانس" یا "مدرک کارشناسی ارشد" است.
- **EducationField**: این ستون رشته تحصیلی یا تخصص تحصیلی کارکنان را مشخص می‌کند. این شامل حوزه‌هایی مانند "بازاریابی"، "مهندسی" یا "مالی" است.
- **Environment Satisfaction**: این ستون میزان رضایت کارکنان را از محیط کاری خود می‌سنجد. که در مقیاسی مانند ۱ تا ۵ نمره گذاری شده است که نشان دهنده سطوح متفاوتی از رضایت است.
- **Gender**: این ستون نشان دهنده جنسیت کارکنان است.
- **JobLevel**: این ستون نشان دهنده سطح سلسله مراتبی یا رتبه پست های کارکنان در داخل شرکت است. می‌توان آن را به صورت عددی نشان داد، مانند ۱، ۲، ۳ و ...
- **MaritalStatus**: این ستون وضعیت تأهل کارمندان را نشان می‌دهد، مانند «مجرد»، «متاهل» یا «طلاق».
- **MonthlyRate**: این ستون به نرخ ماهانه یا حقوق کارکنان اشاره دارد.
- **NumCompaniesWorked**: این ستون نشان دهنده تعداد شرکت‌هایی است که کارمندان قبل از پیوستن به شرکت فعلی در آنها کار کرده‌اند.
- **OverTime**: این ستون نشان می‌دهد که آیا کارمندان اضافه کار می‌کنند یا خیر.
- **PerformanceRating**: این ستون نشان دهنده رتبه بندی یا ارزیابی عملکرد کارکنان است.
- **WorkLife Balance**: این ستون تعادل درک شده کار و زندگی کارکنان را اندازه گیری می‌کند. آن را در مقیاس ۱ تا ۴ نمره داده‌اند که سطوح مختلف تعادل کار و زندگی را نشان می‌دهد.
- **YearsAtCompany**: این ستون نشان دهنده تعداد سال‌هایی است که کارکنان در شرکت فعلی کار کرده‌اند.



یادگیری ماشین

تمرین «یک»

دستیاران آموزشی

روشنک طالشی

امیرمحمد کوشش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۲۹ مهر ۱۴۰۲

۲. [PRISM - ۴۰٪] هدف از این تمرین پیاده‌سازی الگوریتم PRISM است.

الگوریتم PRISM یک الگوریتم است که در داده‌کاوی و تحلیل داده مورد استفاده قرار می‌گیرد. PRISM برای کشف قوانین قابل توجه و سودآور بر اساس استخراج زیرگروه‌های جالب از داده‌ها استفاده می‌کند.

الگوریتم PRISM با استفاده از معیارهایی مانند پوشش^۱ و اطمینان^۲ به دنبال زیرگروه‌هایی در داده‌ها می‌گردد که دارای خواص موردنظر هستند. این الگوریتم با بهره‌گیری از روش‌های تجزیه و تحلیل آماری و ارزیابی جوانب مختلف قوانین، تلاش می‌کند قوانینی را با پوشش بالا و اطمینان قابل قبول پیدا کند.

▪ پیش‌پردازش داده

قبل از اعمال الگوریتم PRISM بر روی داده‌های قلب و عروق cardio.csv، نیاز است داده‌ها را در بخش پیش‌پردازش تمیز کنید. این عمل شامل مراحل زیر می‌شود:

- **حذف داده‌های تکراری:** در صورت وجود داده‌های تکراری، آن‌ها را حذف کنید تا داده‌ها به صورت منحصر به فرد باقی بمانند.
- **پر کردن مقادیر از دست رفته:** در صورتی که برخی از ستون‌ها دارای مقادیر از بین رفته^۳ باشند، می‌توانید این مقادیر را با روش‌هایی مانند میانگین، مد، یا مقدار پرتکرار در آن ستون جایگزین کنید.
- **تبدیل داده‌های عددی به اسمی:** در صورتی که داده‌ها شامل متغیرهای عددی باشند، نیاز است آن‌ها را به صورت اسمی^۴ تبدیل کنید.

در پایان سوال توضیحاتی در مورد ساختار دادگان ارائه شده آورده شده است.

▪ پیاده‌سازی الگوریتم Prism

در این بخش مانند سوال قبل، خروجی قابل انتظار کلاس Prism خواهد بود که دارای حداقل دو متد fit و predict و دیگر متدهای مورد نیاز است. همانطور که پیش‌تر توضیح دادیم در بخش پیش‌پردازش، داده را به دو بخش آموزش و تست به نسبت ۸۰ به ۲۰ به صورت رندوم تقسیم کنید.

^۱ Coverage

^۲ Confidence

^۳ Missing Value

^۴ Categorical



یادگیری ماشین

تمرین «یک»

دستیاران آموزشی

روشنک طالشی

امیرمحمد کوشش پور

دکتر سامان هراتی زاده

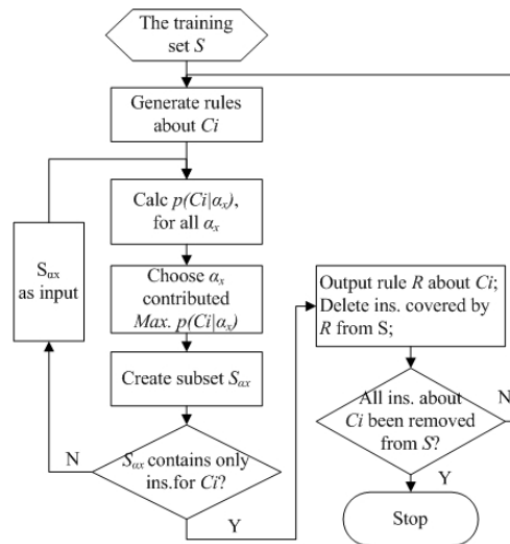
دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۲۹ مهر ۱۴۰۲

متد `fit`: در این بخش، قاعده‌ها بر اساس داده‌های آموزشی تولید می‌شوند. الگوریتم سعی می‌کند با تحلیل داده‌های آموزشی قواعدی را تعیین کند که بر اساس آن، دسته‌بندی موثری بر روی داده‌ها انجام شود. ورودی الگوریتم ماتریس مشخصه `X_train` و بردار هدف `y_train` و خروجی الگوریتم قواعد تولید شده خواهد بود.

متد `predict`: در این متد، داده‌های تست با استفاده از قاعده‌های تولید شده در متد `fit`، دسته‌بندی می‌شوند. این تابع وظیفه پیش‌بینی دسته‌بندی داده‌های تست را بر عهده دارد. در ورودی، داده تست و قاعده را بدهید و در خروجی امتیازهای خواسته شده اعلام شود.



شکل ۲: فلوچارت مربوط به الگوریتم prism

ارزیابی

در ادامه ماتریس درهم‌ریختگی را رسم کنید و نهایتاً با استفاده از معیارهای `precision`، `recall`، `F-score` و `ROC` عملکرد الگوریتم Prism بر روی داده‌های آزمایش بررسی شود. الگوریتم OneRule و Prism را مقایسه کنید در چه شرایطی هر یک از این دو الگوریتم را انتخاب می‌کنید، در گزارش شرح دهید.

ستون‌های دادگان قلبی به شرح زیر است

- **Age**: این ستون نشان دهنده سن افراد در مجموعه داده است.
- **gender**: این ستون جنسیت افراد را نشان می‌دهد. دارای مقادیری مانند "مذکر" و "مونث" باشد که نشان دهنده جنسیت شرکت کنندگان است.



یادگیری ماشین

تمرین «یک»

دستیاران آموزشی
روشنک طالشی
امیرمحمد کوشش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۲۹ مهر ۱۴۰۲

- **height**: این ستون نشان دهنده قد افراد است که معمولاً در سانتی متر اندازه گیری می شود
- **weight**: این ستون نشان دهنده وزن افراد است که معمولاً بر حسب کیلوگرم اندازه گیری می شود.
- **ap_hi**: این ستون نشان دهنده فشار خون سیستولیک افراد است. فشار خون سیستولیک مقدار بالاتری است که در طول اندازه گیری فشار خون اندازه گیری می شود و فشار در شریان ها را در هنگام انقباض قلب منعکس می کند.
- **ap_lo**: این ستون نشان دهنده فشار خون دیاستولیک افراد است. فشار خون دیاستولیک مقدار پایین تری است که در حین اندازه گیری فشار خون اندازه گیری می شود و فشار در شریان ها را زمانی که قلب بین ضربان ها استراحت می کند، منعکس می کند.
- **cholesterol**: این ستون میزان کلسترول افراد را نشان می دهد. دارای دسته‌هایی مانند «طبیعی»، «بیش از حد طبیعی» یا «بالا» باشد که سطوح مختلف کلسترول را نشان می دهد.
- **gluc**: این ستون نشان دهنده سطح گلوکز افراد است. این می تواند دارای دسته‌هایی مانند "طبیعی"، "بالاتر از نرمال" یا "بالا" باشد که سطوح مختلف گلوکز در خون را نشان می دهد.
- **smoke**: این ستون نشان می دهد که افراد سیگاری یا غیرسیگاری هستند.
- **alco**: این ستون نشان می دهد که آیا افراد الکل مصرف می کنند یا خیر.
- **cardio**: این ستون نشان دهنده وجود یا عدم وجود بیماری قلبی عروقی در افراد است. مقادیر دودویی مانند "۱" و "۰" دارد که "۱" نشان دهنده وجود بیماری قلبی عروقی و "۰" نشان دهنده عدم وجود آن است. که این ستون همان کلاس مد نظر ماست.

۳. [۳۰٪ - KNN] هدف از این تمرین پیاده‌سازی الگوریتم نزدیک‌ترین همسایه است.

در این بخش شما می‌بایست به طراحی و پیاده‌سازی طبقه‌بند k -نزدیک‌ترین همسایه^۱ بپردازید، همچنین دادگان ارقام دست‌نویس دودویی^۲ در اختیار شما قرار گرفته است. مجموعه داده ارائه شده شامل اعداد ۰ تا ۹ بوده که در قالب مجموعه‌ای از صفرها و یک‌ها در ابعاد ۲۸ در ۲۸ نمایش داده می‌شوند.

^۱ KNN Classifier

^۲ Digit-recognizer



یادگیری ماشین

تمرین «یک»

دستیاران آموزشی

روشنک طالشی

امیرمحمد کویش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

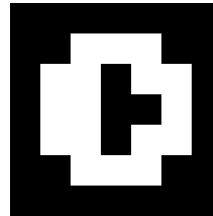
نیم‌سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۲۹ مهر ۱۴۰۲

مجموعه دادگان

هر نمونه در مجموعه دادگان دارای ارتفاعی با ۲۸ پیکسل و عرضی با ۲۸ پیکسل، که در مجموع ۷۸۴ پیکسل است. هر پیکسل می‌تواند دو مقدار صفر و یک را بگیرد، شکلی که پیکسل‌های دارای مقدار «۱» نشان می‌دهند، همان عددی است که توسط کاربر نوشته شده است و بدین صورت مجموعه دادگان آموزشی، دارای ۷۸۵ ستون است، که ستون برچسب^۱ رقمی است که توسط کاربر ترسیم شده است.

```
0000000
0011100
0110110
0110010
0110110
0011100
0000000
```



شکل ۳: نمایش عدد صفر در مجموعه دادگان

پیاده‌سازی

- کلاس KNN که پیاده‌سازی می‌کنید، باید حداقل دارای متدهای زیر باشد.
- 1. `__init__(self, k=3)`
طبقه‌بند KNN را با مقدار پیش‌فرض K (تعداد همسایگان) به صورت ۳ مقداردهی کنید.
- 2. `fit(self, X_train, y_train)`
این متد مدل KNN را بر ماتریس مشخصه `X_train` و بردار هدف `y_train` آموزش می‌دهد.
- 3. `predict(self, X_test)`
این متد داده‌های آزمون `X_test` را می‌پذیرد و برچسب‌های پیش‌بینی شده را بر می‌گرداند.

هر گونه متد یا کلاس اضافی را که برای راه‌حل خود ضروری می‌دانید، اضافه کنید (هدف و استفاده از هر جزء اضافی را مستند کنید).

گزارش تفصیلی

- در قالب PDF رویکرد خود، تصمیمات طراحی، چالش‌های پیش‌روی و راه‌حل اجرا شده را گزارش کنید.
- تاثیر مقادیر مختلف k بر عملکرد مدل را مورد بحث قرار دهید.

¹ Label



یادگیری ماشین

تمرین «یک»

دستیاران آموزشی

روشنک طالشی

امیرمحمد کوشی‌پور

دکتر سامان هراتی‌زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۲۹ مهر ۱۴۰۲

■ برای توضیح منطق و هدف از هر بخش در کد خود، کامنت‌های معناداری^۱ ارائه دهید.

عملکرد شما بر اساس معیارهای زیر ارزیابی خواهد شد؛

- صحت و عملکرد طبقه‌بند KNN.
- شفافیت و سازماندهی کد.
- کیفیت و کامل بودن گزارش.
- نشان‌دادن درک تاثیر پارامترهای کلیدی، مانند K در مدل خود.

بخش امتیازی (+18%)

- دادگان ارائه شده را به تصاویر متناظر تبدیل کنید.
- پیاده‌سازی‌ای برای کلاس KNN ارائه کنید که هم از تصاویر و هم از دادگان متنی برای آموزش و پیش‌بینی استفاده کند.
- روش‌های موجود برای ادغام اطلاعات^۲ از هر دو وجه^۳ برای بهبود دقت را کاوش کنید.
- نه تنها جزئیات فنی، بلکه بینش‌های بکار رفته برای طراحی فرایندهای خود را گزارش کنید.
- در مورد چالش‌های موجود در ترکیب داده‌های چندوجهی و راه‌حل‌های نوآورانه خود بحث کنید.
- نتایج را با تجزیه و تحلیل کمی و کیفی ارائه کنید.

¹ Meaningful Comments

² Information Fusion

³ Modality