



سیستم های عامل

مهلت تحویل:



دانشکده مهندسی برق و  
کامپیوتر

طراحان تمرین:

1403/02/01

استاد:

امیر فراهانی، حامد میرامیرخانی

دکتر مهدی کارگهی

## مقدمه

در این پروژه می‌خواهیم با نحوه مدیریت کردن پردازش‌ها<sup>۱</sup> و روشهای ارتباطی آنها آشنا شویم. در این تمرین، به شبیه‌سازی یکی از روش‌های اساسی و مهم در حوزه یادگیری ماشین<sup>۲</sup> پرداخته می‌شود. یادگیری ماشین یکی از زیرشاخه‌های اصلی هوش مصنوعی است که به تنظیم و اکتشاف روش‌ها و الگوریتم‌هایی می‌پردازد که به کمک آنها، رایانه‌ها و سیستم‌ها می‌توانند اطلاعات را فرا بگیرند و قابلیت پیش‌بینی و تصمیم‌گیری را داشته باشند.

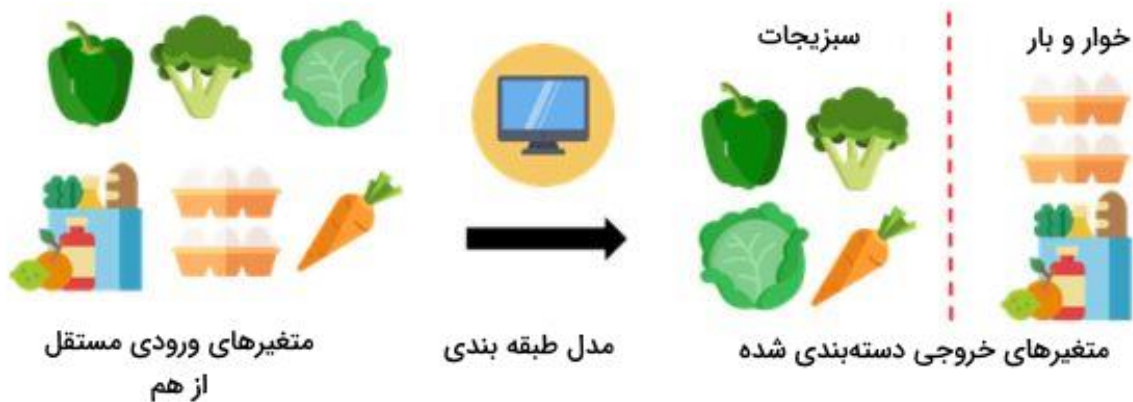
## طبقه‌بندی<sup>۳</sup>

در زمینه یادگیری ماشین طبقه‌بندی به عنوان یکی از مهمترین وظایف محسوب می‌شود که در آن مجموعه‌ای از داده‌ها برای آموزش در دسترس است. در طبقه‌بندی هدف شناسایی تعلق مشاهده جدید به یکی از طبقه‌ها بر اساس مشاهدات قبلی است.

<sup>1</sup> Process

<sup>2</sup> Machine Learning

<sup>3</sup> Classification



#### طبقه بندی خطی<sup>۴</sup>

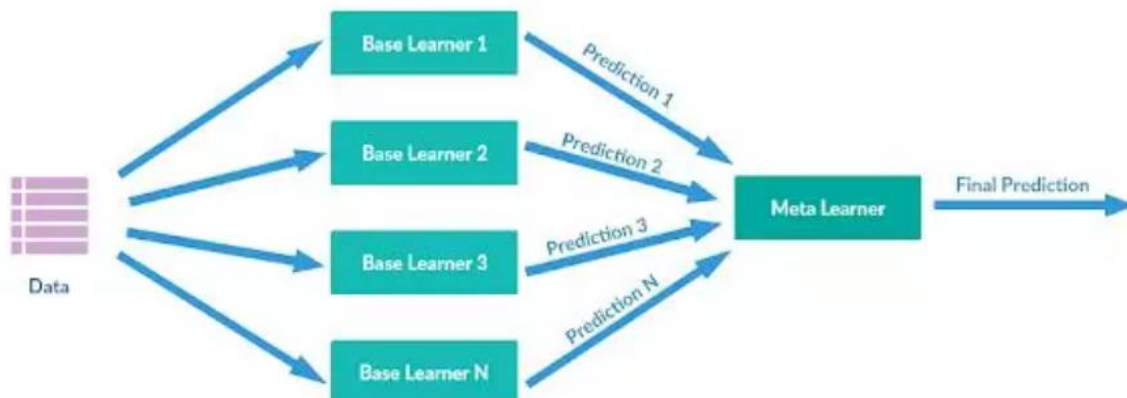
در حوزه یادگیری ماشین، ما با داده هایی سر و کار داریم که می خواهیم ویژگی ها یا طبقه های آن ها را پیش بینی کنیم. این داده ها با استفاده از ویژگی هایی که می توانیم اندازه گیری کنیم و عددی هستند، توصیف می شوند.

الگوریتم های زیادی وجود دارند که با استفاده از یک تابع خطی<sup>۵</sup>، به هر طبقه امتیاز می دهند. این امتیازدهی از طریق ضرب داخلی ویژگی های داده با وزن هر طبقه انجام می شود. طبقه ای که بیشترین امتیاز را دارد، به عنوان طبقه پیش بینی شده انتخاب می شود. فرض کنید می خواهیم با استفاده از اندازه و وزن میوه، تشخیص دهیم آیا آن میوه سیب است یا نه. در اینجا، هر میوه با یک بردار ویژگی معرفی می شود که شامل ویژگی هایی مانند اندازه و وزن است. سپس، با استفاده از ضرب داخلی این بردار و وزن های مربوط به هر طبقه، امتیازی برای هر طبقه محاسبه می شود و سپس طبقه با بیشترین امتیاز به عنوان طبقه پیش بینی شده انتخاب می شود.

<sup>4</sup> Linear classification

<sup>5</sup> Linear Function

## طبقه‌بندی ترکیبی<sup>6</sup>



در حوزه یادگیری ماشین، یک روش پیشرفته برای طبقه‌بندی داده‌ها به نام طبقه‌بندی ترکیبی وجود دارد. این روش به این صورت عمل می‌کند که از چندین طبقه‌بند استفاده می‌کند و نتایج حاصل از هر یک را مورد ترکیب قرار داده و بر اساس بیشترین تکرار، طبقه‌ی نهایی را تعیین می‌کند.

فرض کنید که ما به دنبال تعیین طبقه‌ای برای یک نمونه خاص هستیم. ابتدا این نمونه را به هر یک از طبقه‌بندها که هر کدام دارای بردار وزن مخصوص به خود هستند، می‌دهیم. هر طبقه‌بند بر اساس ویژگی‌های نمونه و با استفاده از بردار وزن خود، نتیجه‌ای را تولید می‌کند و آن را به نمونه تخصیص می‌دهد. سپس از یک رأی‌دهنده<sup>7</sup> استفاده می‌شود که خروجی‌های حاصل از طبقه‌بندها را دریافت کرده و بر اساس تعداد تکرار هر طبقه، طبقه‌ی نهایی را تعیین می‌کند. به عبارت دیگر، طبقه‌ای که بیشترین تعداد بار توسط طبقه‌بندها انتخاب شده است، به عنوان طبقه‌ی نهایی برای نمونه مورد نظر انتخاب می‌شود.

در جلسه توجیهی مثالی درباره این مورد حل میشود. توصیه میشود حتما در این جلسه شرکت کنید.

<sup>6</sup> Ensemble Classification

<sup>7</sup> voter

## شرح تمرین

در این تمرین به شبیه سازی یک طبقه‌بند ترکیبی می‌پردازید. این طبقه بند شامل یک طبقه بند خطی می‌باشد که بردار وزن آن در فایلی جداگانه در اختیار شما قرار داده شده است. وظیفه ای که برنامه شما بر عهده دارد، پیش بینی طبقه مربوط به نمونه هایی می‌باشد که تحت عنوان مجموعه داده اعتبار سنجی در اختیار شما قرار داده شده است.

## ساختار پروژه

برای این پروژه سه نوع پردازش در نظر گرفته شده است:

### الف) پردازش طبقه‌بند ترکیبی

در این فرآیند یک پردازش والد<sup>8</sup> به عنوان سامانه اصلی محسوب میشود که مسئولیت به وجود آوردن پردازش‌های دسته‌بند خطی و معرفی نام پرونده مربوط به بردار وزن دسته‌بند و همینطور به صورت ترتیبی ایندکس فایل بردار فعلی را از طریق یک *Unnamed Pipe* دارد. علاوه بر آن این پردازش وظیفه به وجود آوردن پردازش رأی‌دهنده را نیز بر عهده دارد. پردازش ترکیبی و پردازش رأی‌دهنده از طریق یک *named Pipe* با هم در ارتباطند.

### ب) پردازش دسته‌بند خطی

این پردازش، پردازش فرزند<sup>9</sup> برای پردازش طبقه‌بند ترکیبی محسوب میشود که پس از دریافت نام پرونده بردارهای وزن خود، طبقه مربوط به هر نمونه ای که از طرف طبقه‌بند ترکیبی دریافت میشود را با محاسبه ضرب داخلی بردارهای وزن خود با بردار ویژگی مربوط به نمونه محاسبه کرده و از طریق یک *Unnamed Pipe* به پردازش رأی‌دهنده می‌دهد. (توجه کنید حتما از *Unnamed Pipe* استفاده شود).

### ج) پردازش رأی‌دهنده

---

<sup>8</sup> Parent

<sup>9</sup> Child

این پردازش پس از اتمام هر عملیات طبقه‌بند خطی بر روی مجموعه داده‌های اعتبارسنجی، به پردازش ترکیبی از طریق *named Pipe* اعلام می‌کند تا ایندکس بعدی وارد پردازش خطی شود. در ادامه طبقه‌بند ترکیبی در صورتی که تمام داده‌های ورودی را به طبقه‌بند خطی داده بود از پردازش رای‌دهنده می‌خواهد صحت عملکرد فرآیند را با برگرداندن تعداد پیش‌بینی‌های درست و نادرست به او اطلاع بدهد.

## ورودی و خروجی برنامه

پردازش دسته‌بند ترکیبی، در قالب زیر آدرس مربوط به پوشه<sup>۱۰</sup> بردارهای وزن طبقه‌بندها و پوشه مربوط به داده‌های اعتبارسنجی را از طریق آرگومان<sup>۱۱</sup>‌هایی در رابط خط فرمان<sup>۱۲</sup> از کاربر دریافت می‌کند. برنامه شما باید به ترتیب تعداد پیش‌بینی‌های درست و نادرست را نمایش دهد.

### • نمونه ورودی

```
./main.out rsc/validation rsc/weights
```

### • نمونه خروجی

```
15 3
```

## نکات تکمیلی

- داده‌های اعتبارسنجی در فولدری به نام `validation` قرار داده شده‌اند. در این فولدر فایل‌های به نام `dataset.csv` وجود دارد که مجموعه داده‌های اعتبارسنجی می‌باشد. برچسب‌های مربوط به هر یک از نمونه‌های موجود در این فایل در فایل‌های به نام `labels.csv` در اختیار شما قرار داده شده است.

---

<sup>10</sup> Directory

<sup>11</sup> Argument

<sup>12</sup> Command Line Interface

- بردار وزن مربوط به طبقه بند در فولدری به نام `weights` قرار داده شده است.
- دقت کنید در پایان برنامه باید هر یک از `pipe` ها را `close` کرده باشید و همه ی پردازش ها نیز باید بسته شده باشند.
- دقت کنید در صورتی که علاوه بر موارد ذکر شده در شرح تمرین نیاز به ارسال اطلاعات بیشتری میان پردازش ها بود، ارتباط میان پردازش ها فقط از طریق `pipe` صورت می گیرد و روش های دیگر ارتباط میان پردازش ها قابل قبول نمی باشد.
- در صورتی که تعداد رأی های مربوط به چند طبقه با یکدیگر برابر شد طبقه ای به عنوان طبقه نهایی انتخاب می شود که شماره کوچکتری را دارد.
- برای تجزیه<sup>۱۳</sup> فایل های `CSV` میتوانید از کتابخانه های رایج `C++` برای تجزیه فایل هایی که در اختیارتان قرار گرفته است استفاده کنید.

## نحوه ی تحویل

- برنامه شما باید در سیستم عامل لینوکس و با مترجم `g++` با استاندارد 11 ترجمه و در زمان معقول برای ورودی های آزمون اجرا شود.
- درستی برنامه ی شما از طریق آزمون های خودکار سنجیده می شود. دقت شود که نام فایل اجرایی شما باید `main.out` باشد.
- هدف این تمرین یادگیری شماست. لطفاً تمرین را خودتان انجام دهید در صورت کشف تقلب مطابق قوانین درس با آن برخورد خواهد شد.