

تکلیف سوم

مجتبی ملائی
۴۰۱۳۱۳۸۳

۱. یکی از چالش‌های اصلی در استفاده از شبکه‌های عصبی معمولی برای پردازش تصاویر، تعداد بسیار زیاد پارامترهاست. در این نوع شبکه‌ها، هر نورون به صورت کامل به تمام نورون‌های لایه قبلی متصل است، که در مورد تصاویر با ابعاد بالا، منجر به تولید میلیون‌ها پارامتر می‌شود. به عنوان مثال، یک تصویر رنگی با ابعاد 224×224 دارای بیش از ۱۵۰ هزار ورودی است، که اتصال آن به حتی یک لایه پنهان متوسط باعث افزایش نمایی تعداد وزن‌ها می‌شود. این موضوع نه تنها بار محاسباتی را به طرز چشمگیری افزایش می‌دهد، بلکه باعث دشواری در آموزش شبکه، نیاز به منابع پردازشی قوی، و افزایش رسیک Overfitting می‌شود.

۲. دومین مشکل اساسی شبکه‌های عصبی معمولی در پردازش تصاویر، عدم توانایی در بهره‌برداری از ساختار فضایی و ویژگی‌های محل تصویر است. این شبکه‌ها داده‌های ورودی را به صورت یک بردار یک بعدی در نظر می‌گیرند و هیچ توجهی به موقعیت مکانی پیکسل‌ها یا الگوهای محلی (نظیر لبه‌ها و بافت‌ها) ندارند. در نتیجه، اطلاعات مکانی که برای تحلیل تصویر بسیار حیاتی هستند، نادیده گرفته می‌شوند. این محدودیت باعث می‌شود که مدل نتواند به درستی ویژگی‌های بصری را استخراج کرده و درک عمیقی از محتوای تصویر به دست آورد، در حالی که شبکه‌های کانولوشنی با بهره‌گیری از فیلترهای محلی و اشتراک وزن، این ساختارهای فضایی را به شکل موثرتری مدل‌سازی می‌کنند.

در بینایی کامپیوتر کلاسیک، ویژگی‌ها به صورت دستی و با استفاده از الگوریتم هاو یا فیلترهای از پیش تعریف شده استخراج می‌شوند. این ویژگی‌ها بر اساس دانش انسان و تجربه طراحی می‌شوند و سپس به یک مدل یادگیرنده ساده برای طبقه‌بندی یا شناسایی داده می‌شوند. این فرآیند نیازمند تخصص انسانی و تنظیمات دقیق است و در مواجهه با داده‌های متتنوع یا نویزدار ممکن است عملکرد ضعیفی داشته باشد. در مقابل، روش‌های CNN ویژگی‌ها را به صورت خودکار و سلسه‌مراتبی از داده خام یاد می‌گیرند. شبکه با یادگیری از داده‌های برچسب دار، فیلترهایی را آموزش می‌بیند که به صورت لایه‌به‌لایه از ویژگی‌های ساده (مانند لبه‌ها) تا ویژگی‌های پیچیده (مانند اشیاء یا چهره‌ها) را استخراج می‌کند. این روش نه تنها نیاز به طراحی دستی ویژگی‌ها را حذف می‌کند، بلکه باعث افزایش دقت، انعطاف‌پذیری و عملکرد در ظایف مختلف بینایی کامپیوتر شده است.

در واقع در روش‌های CNN تنظیم این فیلترهاتوسط یادگیری خود مدل تعیین می‌شوند نه براساس طراحی انسان‌ها.

۳

استفاده از Max Pooling به جای Average Pooling در شبکه‌های عصبی کانولوشنی معمولاً ترجیح داده می‌شود، زیرا ویژگی‌های بر جسته‌تر و مهم‌تر تصویر را بهتر حفظ می‌کند و در عمل به بهبود عملکرد مدل در ظایف بینایی کامپیوتر منجر می‌شود. در Max, Pooling بیشترین مقدار در یک ناحیه مشخص انتخاب می‌شود، که به معنای حفظ قوی‌ترین فعال‌سازی‌ها (activations) است. این فعال‌سازی‌ها معمولاً نشان‌دهنده وجود ویژگی‌های مهم مانند لبه‌ها، گوش‌ها یا بافت‌های خاص هستند. در مقابل، میانگین مقادیر را محاسبه می‌کند که می‌تواند اطلاعات کلیدی را با مقادیر غیرمهم ترکیب کرده و در نتیجه ویژگی‌های بر جسته را "تضعیف" کند. بنابراین، Max Pooling با تمرکز روی بر جسته‌ترین نشانه‌ها در تصویر، به استخراج ویژگی‌هایی کمک می‌کند که برای تشخیص و طبقه‌بندی دقیق‌تر حیاتی هستند.

در واقع، Max Pooling همچنین نوعی مقاومت در برابر نویز ایجاد می‌کند؛ زیرا با نادیده گرفتن مقادیر ضعیفتر، تأثیر ناهنجاری‌های کوچک یا تغییرات جزئی در تصویر کاهش می‌یابد. در مجموع، Max Pooling باعث حفظ اطلاعات مهم، کاهش ابعاد، افزایش کارایی محاسباتی، و بهبود قابلیت تعمیم شبکه می‌شود، که دلایل اصلی ترجیح آن بر Average Pooling در اکثر کاربردهای عملی است.

سایز خروجی لایه کانولوشنی:

$$\begin{aligned}n_h[l] &= \lfloor \frac{n_h[l-1] - 2p - f_h}{s} \rfloor + 1 \\n_w[l] &= \lfloor \frac{n_w[l-1] - 2p - f_w}{s} \rfloor + 1 \\n_c[l] &= K\end{aligned}$$

تعداد پارامتر های فیلتر های کانولوشنی:

$$(n_h[l] * n_w[l] * n_c[l] + 1) * K$$

سایز خروجی لایه های پولینگ:

$$\begin{aligned}n_h[l] &= \lfloor \frac{n_h[l-1] - f_h}{s} \rfloor + 1 \\n_w[l] &= \lfloor \frac{n_w[l-1] - f_w}{s} \rfloor + 1 \\n_c[l] &= n_c[l-1]\end{aligned}$$

تعداد پارامتر های FC:

$$(1 + n[l-1]) * n[l]$$

۱. لایه کانولوشنی اول:

$$\begin{aligned}n &= \lfloor \frac{28 - 2 \times 0 - 5}{1} \rfloor + 1 = 24 \\size &= 24 \times 24 \times 6 \\params &= (5 \times 5 \times 1 + 1) \times 6 = 156\end{aligned}$$

۲. لایه avg pooling اول:

$$\begin{aligned}n &= \lfloor \frac{24 - 2}{2} \rfloor + 1 = 12 \\size &= 12 \times 12 \times 6 \\params &= 0\end{aligned}$$

۳. لایه کانولوشنی دوم:

$$\begin{aligned}n &= \lfloor \frac{12 - 2 \times 0 - 5}{1} \rfloor + 1 = 8 \\size &= 8 \times 8 \times 16 \\params &= (5 \times 5 \times 6 + 1) \times 16 = 2416\end{aligned}$$

۴. لایه avg pooling دوم:

$$\begin{aligned}n &= \lfloor \frac{8 - 2}{2} \rfloor + 1 = 4 \\size &= 4 \times 4 \times 16 = 256 \\params &= 0\end{aligned}$$

:FC1 لایه ۵

$$\begin{aligned} \text{input} &= 4 \times 4 \times 16 = 256 \\ \text{size} &= 128 \\ \text{params} &= (256 + 1) \times 128 = 32,896 \end{aligned}$$

:FC2 لایه ۶

$$\begin{aligned} \text{input} &= 128 \\ \text{size} &= 64 \\ \text{params} &= (128 + 1) \times 64 = 8,256 \end{aligned}$$

: softmax لایه خروجی ۷

$$\begin{aligned} \text{input} &= 64 \\ \text{size} &= 10 \\ \text{params} &= (64 + 1) \times 10 = 650 \end{aligned}$$

: مجموع پارامترها ۸

$$\begin{aligned} \text{layers Conv} &= 156 + 2416 = 2,572 \\ \text{layers FC} &= 32,896 + 8,256 + 650 = 41,802 \\ \text{Total} &= 44,374 \end{aligned}$$

۹

خیر، نمونه‌برداری تصادفی ساده از دیتاست برای تقسیم آن به زیرمجموعه‌های آموزش، ارزیابی و آزمایش، همواره کفایت نمی‌کند؛ زیرا در بسیاری از موارد ممکن است این روش منجر به تقسیم نامتوانن یا نامناسب داده‌ها شود که بر عملکرد مدل اثر منفی می‌گذارد. در شرایطی که توزیع کلاس‌ها نامتوانن است (یعنی برخی کلاس‌ها نمونه‌های بسیار بیشتری نسبت به سایر کلاس‌ها دارند)، نمونه‌برداری تصادفی ممکن است باعث شود که برخی کلاس‌ها در یکی از زیرمجموعه‌ها (مثلاً تست یا validation به درستی نمایندگی نشوند یا حتی کاملاً حذف شوند. این امر باعث می‌شود مدل نتواند به درستی یاد بگیرد یا ارزیابی دقیق انجام دهد. به همین دلیل، معمولاً از روش‌های دقیق‌تری مانند نمونه‌برداری طبقه‌بندی شده (stratified sampling) استفاده می‌شود تا نسبت کلاس‌ها در هر زیرمجموعه حفظ شود. همچنین در برخی کاربردها (مثلاً تشخیص چهره یا داده‌های زمانی)، لازم است نمونه‌های وابسته به هم (مانند فریم‌های یک ویدئو یا تصاویر یک فرد) در یک زیرمجموعه باقی بمانند تا از نشت اطلاعات (data leakage) جلوگیری شود. در این موارد نیز نمونه‌برداری تصادفی ساده ناکافی و حتی خطرناک است. در نتیجه، تقسیم‌بندی مؤثر دیتاست نیازمند توجه به ساختار داده، توزیع کلاس‌ها و هدف مدل‌سازی است و صرفاً تکیه بر تصادفی‌سازی کافی نیست.

۱۰

خیر، اضافه کردن یک مقدار ثابت به تمام ورودی‌هایتابع Softmax خروجی آن را تغییر نمی‌دهد. دلیل این موضوع به ویژگی نرم‌السازی نسبی در تابع Softmax بازمی‌گردد. اثبات: برای یک بردار ورودی $\mathbf{z} = [z_1, z_2, \dots, z_n]$ ، خروجی Softmax به صورت زیر تعریف می‌شود:

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}}$$

افزودن مقدار ثابت c به همه ورودی‌ها:

فرض کنید بردار جدید $\mathbf{z}' = [z_1 + c, z_2 + c, \dots, z_n + c]$ باشد. حال خروجی Softmax برای $z_i + c$ برابر است با:

$$\text{Softmax}(z_i + c) = \frac{e^{z_i + c}}{\sum_{j=1}^n e^{z_j + c}} = \frac{e^c \cdot e^{z_i}}{e^c \cdot \sum_{j=1}^n e^{z_j}} = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}}$$

همان‌طور که مشاهده می‌شود، e^c از صورت و مخرج ساده می‌شود، بنابراین خروجی تغییر نمی‌کند.

توصیه می‌شود که در مدل‌های دارای Dropout از نمودار هزینه (Epoch) بر حسب اپیاک (Loss) باعث تعیین نرخ یادگیری به تنها می‌باشد، زیرا نوسانات ذاتی Dropout باعث ناپایداری در مقدار هزینه می‌شود و ممکن است اطلاعات گمراه‌کننده درباره روند یادگیری ارائه دهد.

در فرآیند Dropout، در هر بار forward pass، زیرمجموعه‌ای تصادفی از نورون‌ها غیرفعال می‌شوند. این امر باعث می‌شود که مقدار تابع هزینه از یک اپیاک به اپیاک دیگر رفتار نوسانی و پراز جهش‌های تصادفی داشته باشد. در نتیجه، نمودار هزینه ممکن است دارای افت و خیزهای زیاد باشد که تشخیص روند کلی همگرایی یا واگرایی مدل را دشوار می‌کند. این امر می‌تواند موجب انتخاب نامناسب نرخ یادگیری شود، چرا که تحلیل‌گر ممکن است نوسانات را به نرخ یادگیری نامطلوب نسبت دهد در حالی که منشأ آن Dropout است.

۱. **کاهش یا تغییر ابعاد کانال‌ها:** کانولوشن 1×1 تنها بر بعد عمقی (تعداد کانال‌ها) اثر می‌گذارد و بدون تغییر در ابعاد مکانی (طول و عرض تصویر)، می‌تواند تعداد کانال‌ها را کاهش یا افزایش دهد. این قابلیت برای کاهش پیچیدگی محاسباتی یا افزایش قدرت بازنمایی در لایه‌های خاص سیار مفید است. به عنوان مثال، در U-Net از کانولوشن 1×1 برای تطبیق تعداد کانال‌ها قبل از اتصال لایه‌های encoder و decoder استفاده می‌شود.

۲. **افزایش غیرخطی بودن و ترکیب ویژگی‌ها:** با اعمال فیلتر 1×1 روی چندین کانال، مدل قادر است بین ویژگی‌های مختلف ترکیب‌های غیرخطی جدیدی یاد بگیرد. این عملیات شبیه به یک لایه Connected Fully معمولاً در روی هر موقعیت مکانی (pixel) است و به مدل اجازه می‌دهد تا وابستگی‌های میان کانالی را در همان مکان بررسی کند، که برای استخراج ویژگی‌های سطح بالا مفید است.

۱.۹ نشانه‌ها

۱. **اختلاف زیاد بین دقت آموزش و دقت اعتبارسنجی:** مدل در داده‌های آموزش عملکرد بسیار خوبی دارد (خطای کم یا دقت بالا)، اما در داده‌های اعتبارسنجی (Validation) عملکرد ضعیفی نشان می‌دهد. این اختلاف نشان‌دهنده یادگیری بیش از حد ویژگی‌های خاص داده‌های آموزش است، نه الگوهای عمومی.

۲. **کاهش پیوسته خطای آموزش، اما ثابت یا افزایش یافتن خطای اعتبارسنجی:** اگر با گذشت اپیاک‌ها، خطای آموزش کاهش یابد ولی خطای اعتبارسنجی بهبود نیابد یا حتی افزایش یابد، این نشانه واضحی از overfitting است.

۲.۹ دو تکنیک کاربردی برای کاهش :Overfitting

۱. **استفاده از Regularization یا Dropout مانند Regularization L2 :**

- Dropout:** به صورت تصادفی برخی نورون‌ها را در طول آموزش غیرفعال می‌کند، که از وابستگی بین از حد مدل به مسیرهای خاص جلوگیری می‌کند. در CNN‌ها معمولاً در لایه‌های Connected Fully استفاده می‌شود.

- Regularization L2:** با اضافه کردن جرمیه به وزن‌های بزرگ، مدل را وادار می‌کند تا پارامترهای ساده‌تر و عمومی‌تری بیاموزد.

۲. **افزایش داده‌ها از طریق Data Augmentation:** با ایجاد تغییرات تصادفی در تصاویر آموزشی (چرخش، برش، مقیاس‌گذاری، وارونه‌سازی، تغییر روش‌نایابی و غیره)، داده‌های متنوع‌تری ایجاد می‌شود که کمک می‌کند مدل ویژگی‌های پایدار و عمومی‌تری یاد بگیرد و از حفظ ویژگی‌های خاص و جزئی تصویر اصلی اجتناب کند.

در مدل‌های یادگیری عمیق، خروجی Softmax معمولاً به عنوان توزیع احتمال روی کلاس‌ها تفسیر می‌شود. اما در عمل، این احتمال‌ها اغلب overconfident هستند، یعنی مدل حتی در شرایط عدم قطعیت، احتمال بسیار بالایی برای یک کلاس خاص اعلام می‌کند. این موضوع باعث کاهش قابلیت اطمینان مدل در سیستم‌های حساس به تصمیم‌گیری می‌شود. برای کالیبره‌سازی (Calibrating) بهتر خروجی، از پارامتر دمایی (Temperature Scaling) استفاده می‌شود. در این روش، خروجی logits مدل (بردار قبل از Softmax) بر یک پارامتر $T < 0$ تقسیم می‌شود:

$$\text{Softmax}_T(z_i) = \frac{e^{z_i/T}}{\sum_j e^{z_j/T}}$$

- وقتی $T = 1$: خروجی معمولی Softmax تولید می‌شود.
- وقتی $T > 1$: توزیع نرم‌تر می‌شود؛ مدل با اطمینان کمتری پیش‌بینی می‌کند.
- وقتی $T < 1$: توزیع تیزتر می‌شود؛ مدل با اطمینان بیشتری تصمیم می‌گیرد.

در مرحله کالیبراسیون، مقدار T روی مجموعه اعتبارسنجی (Validation) بهینه‌سازی می‌شود تا فاصله بین توزیع پیش‌بینی شده و واقعیت آماری (مثلاً با معیارهای مانند Expected Calibration Error) کمینه شود.

مثال کاربردی: سیستم‌های تصمیم‌گیری در پزشکی

در یک سیستم تشخیص پزشکی با استفاده از شبکه عصبی کانولوشنی، خروجی مدل ممکن است احتمال "سرطان بودن" یا "سلام بودن" یک بافت را گزارش دهد. اگر مدل بدون کالیبره‌سازی Softmax مقدار ۹۸.۰٪ برای "سرطان" تولید کند، ولی در واقعیت تنها در ۸۰٪ موارد این پیش‌بینی درست باشد، سیستم تصمیم‌گیر ممکن است به شدت گمراه شود. با اعمال Scaling، Temperature، این احتمال ممکن است به مقدار واقع‌گرایانه‌تری مانند ۸۰.۰٪ تنظیم شود، که منجر به تصمیم‌گیری قابل اعتمادتر توسط پزشک یا سیستم اتوماسیون می‌شود.

۱۱

در معماری‌های بسیار عمیق، CNN دو چالش اصلی وجود دارد: مشکل ناپدید شدن/انفجار گرادیان‌ها (vanishing/exploding gradients) و مشکل افت عملکرد با افزایش عمق شبکه (degradation problem). به طور خلاصه، هنگامی که شبکه‌ها عمیق می‌شوند، گرادیان‌های backpropagation در لایه‌های اولیه به طور تدریجی کاهش یا افزایش می‌یابند که فرآیند یادگیری را نامن و نایاب‌دار می‌سازد. علاوه بر این، افزایش عمق منجر به کاهش عملکرد شبکه در برخی موارد می‌شود چرا که شباهت‌های اطلاعاتی و وابستگی‌های پیچیده بین لایه‌ها به طور ناکارآمد منتقل می‌گردد و مدل نتواند الگوهای کلی را به طور بهینه استخراج کند.

تکنیک‌هایی مانند skip connections در معماری ResNet با ارائه مسیرهای معادل (identity mapping) برای انتقال، این مشکلات را برطرف می‌کنند. این اتصالات با ایجاد ارتباط مستقیم بین لایه‌های اولیه و نهایی شبکه، اجازه می‌دهند تا گرادیان‌ها بدون از دست دادن اطلاعات اصلی به عقب منتقل شوند؛ به عبارت دیگر، مدل قادر است وابستگی‌های عمیق‌تری را بدون رسک از دست دادن اطلاعات مفید، یاد بگیرد. از منظر عملکردی، این ساختار باعث می‌شود که بهبودهای قابل توجهی در همگرایی، دقت و تعمیم‌پذیری مدل حاصل شود، به گونه‌ای که ساختارهای بسیار عمیق بدون مشکلات کاهش عملکرد، به راحتی قابل آموزش باشند.

۱۲ بخش الگوریتم ژنتیک

۱.۱۲

(آ) یک کروموزوم به طول تعداد دانشجویان در نظر می‌گیریم. در این کروموزوم خانه نشان دهنده انتخاب دانشجوی i است و مقدار آن نشان دهنده پروژه آن است.

$$\text{Chromosome} = [P_{j_1}, P_{j_2}, P_{j_3}, \dots, P_{j_s}]$$

به عنوان مثال: Chromosome = [1, 7, 9, 2, 3] یعنی دانشجوی اول پروژه اول، دانشجوی دوم پروژه هفتم و ...

(ب)

| ۱۳ پ | ۱۲ پ | ۱۱ پ | ۱۰ پ | ۹ پ | ۸ پ | ۷ پ | ۶ پ | ۵ پ | ۴ پ | ۳ پ | ۲ پ | ۱ پ | دانشجو |
|------|------|------|------|------|------|------|------|------|------|------|------|------|--------|
| ۵ | ۱۰۰۰ | ۱۰۰۰ | ۴ | ۱۰۰۰ | ۳ | ۱۰۰۰ | ۱۰۰۰ | ۲ | ۱۰۰۰ | ۱۰۰۰ | ۱ | ۱۰۰۰ | ۱ |
| ۱۰۰۰ | ۱۰۰۰ | ۱۰۰۰ | ۵ | ۴ | ۱۰۰۰ | ۱۰۰۰ | ۳ | ۱۰۰۰ | ۱۰۰۰ | ۱۰۰۰ | ۱ | ۲ | ۲ |
| ۱۰۰۰ | ۱۰۰۰ | ۳ | ۱۰۰۰ | ۱۰۰۰ | ۱۰۰۰ | ۵ | ۱۰۰۰ | ۴ | ۱ | ۱۰۰۰ | ۲ | ۱۰۰۰ | ۳ |
| ۱۰۰۰ | ۱۰۰۰ | ۱۰۰۰ | ۴ | ۵ | ۲ | ۱۰۰۰ | ۱ | ۱۰۰۰ | ۱۰۰۰ | ۳ | ۱۰۰۰ | ۱۰۰۰ | ۴ |
| ۵ | ۱۰۰۰ | ۱۰۰۰ | ۱۰۰۰ | ۱۰۰۰ | ۱۰۰۰ | ۲ | ۱۰۰۰ | ۳ | ۱۰۰۰ | ۱۰۰۰ | ۴ | ۱ | ۵ |
| ۱۰۰۰ | ۴ | ۵ | ۱۰۰۰ | ۱ | ۱۰۰۰ | ۱۰۰۰ | ۳ | ۱۰۰۰ | ۲ | ۱۰۰۰ | ۱۰۰۰ | ۱۰۰۰ | ۶ |
| ۱۰۰۰ | ۱۰۰۰ | ۱۰۰۰ | ۲ | ۱۰۰۰ | ۱ | ۵ | ۱۰۰۰ | ۱۰۰۰ | ۱۰۰۰ | ۴ | ۱۰۰۰ | ۳ | ۷ |
| ۱۰۰۰ | ۱۰۰۰ | ۴ | ۱۰۰۰ | ۱۰۰۰ | ۱۰۰۰ | ۱ | ۲ | ۱ | ۱۰۰۰ | ۳ | ۱۰۰۰ | ۱۰۰۰ | ۸ |
| ۳ | ۱۰۰۰ | ۱۰۰۰ | ۱۰۰۰ | ۱۰۰۰ | ۴ | ۲ | ۱ | ۱۰۰۰ | ۱۰۰۰ | ۵ | ۱۰۰۰ | ۱۰۰۰ | ۹ |

جدول ۱: جدول انتخاب دانشجویان

مقدار جریمه در این مثال ۱۰۰۰ در نظر گرفته شده است. جمعیت اولیه

| Chromosome | <i>S</i> ₁ | <i>S</i> ₂ | <i>S</i> ₃ | <i>S</i> ₄ | <i>S</i> ₅ | <i>S</i> ₆ | <i>S</i> ₇ | <i>S</i> ₈ | <i>S</i> ₉ |
|-------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 5 | 3 | 4 | 2 | 7 | 4 | 10 | 6 | 9 |
| 3 | 8 | 9 | 11 | 3 | 5 | 6 | 1 | 2 | 12 |
| 4 | 1 | 6 | 5 | 10 | 2 | 12 | 3 | 11 | 13 |
| 5 | 10 | 12 | 7 | 9 | 13 | 11 | 7 | 4 | 3 |

ج) اگر دو دانشجو یک پروژه را بردارند باید برای آن جرمیه تعريف کنیم. (متلا ۱۰۰۰)

- اگر دانشجویی نیز پروژه ای دارد که در ۵ اولولیت آن نبوده باید برای آن جرمیه تعريف کنیم. (قبل تعريف شده است)
- حداکثر جرمیه ممکن برابر است با وقتی که همه دانشجویان یک پروژه یکسان بردارند که در اولولیت هیچ کدام نیست.
- همچنین باید تلاش شود تا دانشجویان اولولیت های بالا تر خود برسند.
- تابع dupcnt تکراری ها را میشمارد.

تابع fitness زیر اینکار را انجام میدهد:

$$f(c) = \frac{1}{\sum_i r_{ij} + dupcnt(c)*1000} * 9 * 2000$$

این تابع همه شرایط لازم را دارد. و خروجی آن بین ۱ تا ۲۰۰۰ برای این مسئله میباشد.

```

1 student_preferences = [
2   [1000, 1, 1000, 1000, 2, 1000, 1000, 3, 1000, 4, 1000, 1000, 5],
3   [2, 1, 1000, 1000, 1000, 3, 1000, 1000, 4, 5, 1000, 1000, 1000],
4   [1000, 2, 1000, 1, 4, 1000, 5, 1000, 1000, 1000, 3, 1000, 1000],
5   [1000, 1000, 3, 1000, 1000, 1, 1000, 2, 5, 4, 1000, 1000, 1000],
6   [1, 4, 1000, 1000, 3, 1000, 2, 1000, 1000, 1000, 1000, 1000, 5],
7   [1000, 1000, 1000, 2, 1000, 3, 1000, 1000, 1, 1000, 5, 4, 1000],
8   [3, 1000, 4, 1000, 1000, 1000, 5, 1, 1000, 2, 1000, 1000, 1000],
9   [1000, 3, 1000, 5, 1, 2, 1000, 1000, 1000, 1000, 4, 1000, 1000],
10  [1000, 1000, 5, 1000, 1000, 1000, 1, 2, 4, 1000, 1000, 1000, 3]
11 ]
12 chromosomes = [
13   [2, 1, 2, 1, 1, 1, 1, 1, 1, 1],
14   [5, 3, 4, 2, 7, 4, 10, 6, 9],
15   [8, 9, 11, 3, 5, 6, 1, 2, 12],
16   [1, 6, 5, 10, 2, 12, 3, 11, 13],
17   [10, 12, 7, 9, 13, 11, 7, 4, 3]
18 ]
19 def fitness(chromosome, preferences):
20     total_cost = 0
21     duplicate_penalty = 0
22     seen = set()
23
24     for student_index, project in enumerate(chromosome):
25         project_index = project - 1
26         cost = preferences[student_index][project_index]
27         total_cost += cost
28
29         if project in seen:
30             duplicate_penalty += 1
31         else:
32             seen.add(project)
33
34     penalty = duplicate_penalty * 1000
35     adjusted_cost = total_cost + penalty
36
37     fitness_value = (1 / adjusted_cost) * 9 * 2000
38     return fitness_value
39
40 # Calculate fitness for all chromosomes
41 fitness_values = [fitness(ch, student_preferences) for ch in chromosomes]
42 for i, fval in enumerate(fitness_values, start=1):
43     print(f"Fitness of Chromosome {i}: {fval:.4f}")
44
45

```

output:

```
1   Fitness of Chromosome 1: 1.6350
2   Fitness of Chromosome 2: 5.9701
3   Fitness of Chromosome 3: 17.5610
4   Fitness of Chromosome 4: 17.4757
5   Fitness of Chromosome 5: 8.8279
```

$p3 = [8, 9, 11, 3, 5, 6, 1, 2, 12]$
 $p4 = [1, 6, 5, 10, 2, 12, 3, 11, 13]$
 $c1 = [8, 6, 5, 3, 2, 12, 1, 11, 13] \rightarrow 580.6451$
 $c2 = [1, 9, 11, 10, 5, 6, 3, 2, 12] \rightarrow 8.8932$

$p2 = [5, 3, 4, 2, 7, 4, 10, 6, 9]$
 $p3 = [8, 9, 11, 3, 5, 6, 1, 2, 12]$
 $c3 = [8, 3, 11, 2, 5, 4, 7, 6, 9] \rightarrow 8.9020$
 $c4 = [5, 9, 4, 3, 7, 6, 10, 2, 12] \rightarrow 17.6470$

• ترکیب کروموزم سوم و چهارم (d)

اگر از uniform crossover استفاده کنیم:

• ترکیب کروموزم دوم و سوم

اگر از uniform crossover استفاده کنیم:

(e) برگزیدگان:

| | |
|----|----------|
| c1 | 580.6451 |
| c4 | 17.6470 |
| p3 | 17.5610 |
| p4 | 17.4757 |
| c3 | 8.9020 |

برازندگی کل نسل قبلی : 51.4697

برازندگی کل نسل جدید : 642.2308

۲.۱۲

(آ) نمایش مناسب برای کروموزوم‌ها: یک کروموزوم می‌تواند به صورت یک رشته از شناسه‌های مشتریان به همراه شناسه وسیله نقلیه در قالب لیست‌های مجزا نمایش یابد. برای مثال:

$$\text{Chromosome} = [[v_1 : c_3 \rightarrow c_5 \rightarrow c_1], [v_2 : c_2 \rightarrow c_6], [v_3 : c_4]]$$

که در آن v_i نمایانگر وسیله نقلیه و c_j نشان‌دهنده مشتری زام است. این نمایش هم اختصاص مشتریان به وسائل نقلیه و هم ترتیب ملاقات را مشخص می‌کند.

(ب) تعریف اپراتور قیاسی با جریمه تابع هزینه به صورت زیر تعریف می‌شود:

$$f = \sum_{i=1}^{m_1} d_i + \lambda_1 \cdot P_{\text{capacity}} + \lambda_2 \cdot P_{\text{timewindow}}$$

:که

- مسافت طی شده توسط وسیله نقلیه i است.
- مجموع جریمه‌های ناشی از تجاوز ظرفیت وسائل نقلیه است.
- مجموع جریمه‌های ناشی از تخطی پنجره زمانی مشتریان است.
- λ_2 و λ_1 ضرایب جریمه هستند.

(ج) طراحی عملگرهای Mutation و Crossover

:Crossover

استفاده از عملگر Route-Based Crossover (RBX)

- یک یا چند مسیر (route) تصادفی از والد اول انتخاب شده و بدون تغییر به فرزند منتقل می‌شوند.
- مشتریانی که در این مسیرها نیستند، از والد دوم و به ترتیبی که در والد دوم آمده‌اند، در جای خالی باقی‌مانده قرار می‌گیرند به‌گونه‌ای که تکراری نباشند.

:مثال:

Parent 1: $[[v_1 : c_1 \rightarrow c_2], [v_2 : c_3 \rightarrow c_4]]$
 Parent 2: $[[v_1 : c_4 \rightarrow c_1], [v_2 : c_2 \rightarrow c_3]]$

فرض کنید مسیر $[v_1 : c_1 \rightarrow c_2]$ از والد اول انتخاب شود. مشتریان باقی‌مانده (c_3, c_4) طبق ترتیب والد دوم به فرزند افزوده می‌شوند، نتیجه:

Offspring: $[[v_1 : c_1 \rightarrow c_2], [v_2 : c_4 \rightarrow c_3]]$

:Mutation

استفاده از دو عملگر جهش:

- **Mutation: Swap** جابجایی یک مشتری از مسیر فعلی‌اش به موقعیتی دیگر در همان مسیر یا مسیر متفاوت.
- **Mutation: Relocate**

:Swap مثال

$$[v_1 : c_1 \rightarrow c_2 \rightarrow c_3] \Rightarrow [v_1 : c_1 \rightarrow c_3 \rightarrow c_2]$$

:Relocate مثال

$$[v_1 : c_1 \rightarrow c_2], [v_2 : c_3] \Rightarrow [v_1 : c_2], [v_2 : c_3 \rightarrow c_1]$$

(d) انتخاب والد و بازمانده

برای انتخاب والدین از روش (Tournament Selection) استفاده می‌شود:

- کروموزوم به صورت تصادفی انتخاب شده و بهترین آن‌ها بر اساس تابع هزینه به عنوان والد انتخاب می‌شود.

برای انتخاب بازماندان از روش **Elitism** استفاده می‌شود:

- درصدی از بهترین کروموزوم‌های نسل فعلی مستقیماً به نسل بعد منتقل می‌شوند.
- بقیه جمعیت از بین فرزندان تولیدشده انتخاب می‌گردند.

(e) شرط توقف

شرایط توقف می‌تواند یکی از موارد زیر باشد:

- رسیدن به تعداد نسل معین G_{\max} .
- عدم بهبود تابع هزینه برای k نسل متوالی.
- رسیدن به مقدار حداقل مطلوب برای تابع هزینه.