

دانشگاه اصفهان دانشکده مهندسی کامپیوتر

عنوان:

تمرین اول هوش محاسباتی: الگوریتمهای ژنتیک

Genetic Algorithms

نگارش

دانیال شفیعی مهدی مهدیه امیررضا نجفی

استاد راهنما

دكتر كارشناس

درس مبانی هوش محاسباتی صفحه ۲ از ۱۵

لب	مطا	ست	فهر
			<i>_</i> 0

۲	م <i>قد</i> مه	•
۲	مبانی و مفاهیم الگوریتم ژنتیک	١
٣	درک و حل مسائل با الگوریتم ژنتیک	4
٧	پیادهسازی، ارزیابی و تجزیه و تحلیل الگوریتم ژنتیک	٣

• مقدمه

هدف از این تمرین آشنایی بیشتر با الگوریتمهای ژنتیک و استفادهی بیشتر از آنها در کاربردهای عملی است.

۱ مبانی و مفاهیم الگوریتم ژنتیک

۲ درک و حل مسائل با الگوریتم ژنتیک

- ۱. (آ) اگر هیچ گرهای نباید دو بار دیده شود، یک کروموزوم باید باید یک دور بین همهی گرهها باشد که این دور شامل طی ترتیب طی کردن ۱۰ یال است. پس کروموزوم ما شامل ۱۰ ژن است.
- (ب) اینکه بین کدام شهرها ارتباط وجود داشته باشد پیش فرضهای مسئله است اما به طور کلی می توان گفت اگر گراف کامل و بدون طوقه باشد، از هر گرهای به همهی گرههای دیگر یال وجود دارد. ما در نظر گرفتیم این یالها جهت دار $n imes \frac{n-1}{2}$ است پس اگر از یک گره به گره ی دیگر رفت برگشت نیازی نیست. با این اوصاف تعداد کل ژنهای ممکن $n imes \frac{n-1}{2}$ یال می شود. که اینجا n imes 10 است پس n imes 45 اشت وجود دارد.
 - ۲. (آ) ژنها را به تابع fitness میبریم:

$$fit(x_1) = 6 + 5 - 4 - 1 + 3 + 5 - 3 - 2 = 9$$

$$fit(x_2) = 8 + 7 - 1 - 2 + 6 + 6 - 0 - 1 = 23$$

$$fit(x_3) = 2 + 3 - 9 - 2 + 1 + 2 - 8 - 5 = -16$$

$$fit(x_4) = 4 + 1 - 8 - 5 + 2 + 0 - 9 - 4 = -19$$

به ترتیب x_2 به x_3 ، x_1 ، x_2 برازنده هستند.

(ب) عملیات ترکیب

• ترکیب نقطهای: در این روش به دو فرزند جدید میرسیم.

$$x_{21} = 8712|3532$$

$$x_{21} = 6541|6601$$

• ترکیب دو نقطهای: با استفاده از این روش به دو فرزند جدید میرسیم. ما فرض میکنیم منظور از نقاط b و f یعنی بعد از این نقاط ترکیب اتفاق می افتد

$$x_{131} = 65|9212|35$$

$$x_{313} = 23|4135|85$$

• ترکیب یکنواخت: برای انجام این ترکیب نیازمند به یک ماسک هستیم. این ماسک یک ژن تصادفی با مقادیر دودویی است که نشانگر این است که آن ژن را از کروموزوم اول بگیریم یا دوم. که انتخاب اول یا دوم هم احتمال است. ما با استفاده از

برنامهٔ ۱: تولید ماسک تصادفی

```
import random
mask = ''.join(random.choice('01') for _ in range(8))
r print(mask)
```

درس مبانی هوش محاسباتی صفحه ۴ از ۱۵

یک رشته ی تصادفی از ۰ و ۱ تولید می کنیم. ما فرض می کنیم ۰ معادل رشته ی اول و ۱ معادل رشته ی سوم باشد.

mask = 01001010

 $x_{13} = 8|3|12|1|6|8|1$

 $x_{31} = 2|7|92|6|2|0|5$

(ج) برازش فرزندان: با استفاده از تکه کد زیر برازندگی هر فرزند را محاسبه میکنیم:

برنامهٔ ۲: محاسبه ی برازندگی

```
chromosome = input()
a, b, c, d, e, f, g, h = [int(char) for char in chromosome]
fitness = a + b - c - d + e + f - g - h
print(fitness)
```

$$fit(x_{21}) = 87123532 = 15$$

$$fit(x_{21}) = 65416601 = 17$$

$$fit(x_{131}) = 65921235 = -5$$

$$fit(x_{313}) = 23413585 = -5$$

$$fit(x_{23}) = 83121681 = 6$$

$$fit(x_{32}) = 27926205 = 1$$

تعبیر بهتر شدن و بدتر شدن تعبیر نا دقیقی است. ما دو شاخص را برای بهتر شدن و بدتر شدن در نظر میگیریم.

۱. بالاترین برازندگی: در والدها بالاترین برازندگی ۲۳ بود که به ۱۷ کاهش یافت یعنی بدتر شده.

۲. میانگین برازندگی: در شرایط قبلی برازندگی معادل
$$\frac{9+23-16-19}{4}=\frac{-3}{4}=-0.75$$
 میشود و در فرزندان ۲. میانگین برازندگی: در شرایط قبلی برازندگی معادل $\frac{15+17-5-5+6+1}{6}=\frac{29}{6}\approx 4.83$

- (ه) ما سعی کردیم بهترین ترکیب را بسیازیم و آن $x_{\rm optimal} = 87116601$ خواهد بود که برازندگی آن ۲۴ خواهد شد. پس نمی توان بدون جهش به نقطه ی بهینه رسید و حداقل ۱۲ تا فاصله با نقطه ی برازندگی وجود خواهد داشت.

kjhghghghghg . "

(آ) مقدار برازندگی به ازای هر x:

$$fit(x_1) = 1 - 4 + 7 = 4$$

$$fit(x_2) = 8 - 16 + 7 = -1$$

$$fit(x_3) = 27 - 36 + 7 = -2$$

$$fit(x_4) = 64 - 64 + 7 = 7$$

(ب) بله. می توانیم با اضافه کردن $c \geq 2$ همه ی مقدارها را نامنفی کنیم. مثلا اگر c = 3 در نظر بگیریم رابطه ی برازندگی $\operatorname{fit}(x) = x^3 - 4x^2 + 10$ خواهد شد.

(ج) به هر برازندگی مقدار ثابت ۲ اضافه می شود پس

TotalFitness =
$$(4+3) \times 2 + (-1+3) \times 3 + (-2+3) \times 3 + (7+3) \times 2$$

$$= 14 + 6 + 3 + 20 = 43$$

(د) مقدار برازندگی نسبی برای هر نمونه x به صورت زیر خواهد شد:

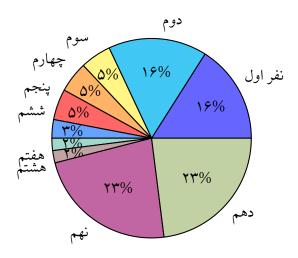
$$P(x=1) = \frac{7}{43} = 0.1628$$

$$P(x=2) = \frac{2}{43} = 0.0465$$

$$P(x=3) = \frac{1}{43} = 0.0233$$

$$P(x=4) = \frac{10}{43} = 0.2326$$

میتوانیم آن را به صورت یک گردونه هم نشان دهیم



شکل ۱: گردونهی شانس برای این نمونه از جمعیت

(ه) مزیت تابع جدید این است که به ازای هر مقدار x، تابع برازندگی همواره نامنفی است. برای محاسبه g(x) تمام

مقدادیر بدست آمده در بخش آ را به توان ۲ میرسانیم.

$$fit(x_1) = 4^2 = 16$$

$$fit(x_2) = (-1)^2 = 1$$

$$fit(x_3) = (-2)^2 = 4$$

$$fit(x_4) = 7^2 = 49$$

- (و) فشار انتخاب: فشار انتخاب یعنی درجه ی اینکه افراد برازنده تر چقدر شانس زنده ماندن دارند. برعکس اضافه کردن مقدار ثابت، در «به توان رساندن» فشار انتخاب زیاد می شود. البته این تا حدودی بستگی به روش انتخاب هم دارد. مثلا اگر از الگوریتم انتخاب رتبه پایه ۱ استفاده کنیم دیگر این مسئله جدی نیست.
- همگرایی: میتوان گفت با افزایش فشار انتخاب، همگرایی سریعتر میشود اما خطر گیر کردن در یک نقطهی بهینهی محلی وجود دارد.
- تنوع: افزایش فشار انتخاب باعث کاهش تنوع و همگرایی سریع به یک قلهی محلی خاص می شود که ممکن است بهترین نباشد. با افزایش فشار انتخاب تنوع در انتخاب گونهها را از دست می دهیم.

rank-based selection'

درس مبانی هوش محاسباتی صفحه ۷ از ۱۵

۳ پیادهسازی، ارزیابی و تجزیه و تحلیل الگوریتم ژنتیک جهت انتخاب بهترین ویژگی برای مسئلهی واقعی دستهبندی مشتریان

۱. پیش پردازش دادهها:

(آ) حذف دادههای پرت:

برای پر کردن دادههای پرت از روش IQR method استفاده میکنیم این روش به این صورت است که IQR را برای پر کردن دادههای کوچکتر از Q1 - Q1 - 1.5*IQR برابر با Q3 - Q1 سپس دادههای کوچکتر از Q3 + Q1 - 1.5*IQR و بزرگتر از Q3 + Q1 - 1.5*IQR را حذف میکنیم.

برنامهٔ ۳: حذف دادههای پرت

```
df = pd.read_csv('Customer Classification dataset/Train.csv')

numeric_cols = df.select_dtypes(include=[np.number]).columns.tolist()

numeric_cols.remove('ID')

for col in numeric_cols:
    Q1 = df[col].quantile(0.25)
    Q3 = df[col].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    df = df[(df[col] >= lower_bound) & (df[col] <= upper_bound)]</pre>
```

(ب) رمزگذاری ویژگیهای دسته ای (categorical):

مهم ترین پارامتر در رمزگذاری (encoding) داده های دسته ای این است که ببینیم که این داده ها داده های کیفی ترتیبی (Ordinal Qualitative) هستند یا کیفی اسمی (Nominal Qualitative) اگر داده ها کیفی ترتیبی باشند نیاز داریم که داده ها را به گونه ای پیش پردازش بکنیم که این ترتیب همچنان حفظ شود و در صورتی که داده ها اسمی باشند نیازی به این کار نداریم و میتوانیم برای هر دسته یک ستون درست کنیم که با مقادیر درست (True) و غلط (False) مشخص کنیم که به این دسته قرار دارد یا خیر.

Profession ، Graduated ، Ever_Married ، Gender در داده هایی که در سوال به ما داده شده بود ستون های Var_1 و Var_1 ستونهایی بودند که داده های انها به صورت کیفی اسمی بود و برای همین برای رمز گذاری انها از var_1 استفاده کردیم.

برنامهٔ ۴: رمزگذاری با get_dummies

درس مبانی هوش محاسباتی صفحه ۸ از ۱۵

```
prefix=columns_to_encode,

drop_first=True

features.head()
```

و ستون Spending_Score را که دارای داده های ترتیبی به نام های Average ، Low و High بود را به صورت دیگری رمزگذاری کردیم تا مدل توانایی درک این که این ۳ مقدار دارای ترتیب مشخصی هستند را متوجه شود. اینکار را به این صورت انجام دادیم که این ۳ ستون را به ترتیب از کوچک به بزرگ از ۰ تا ۲ مقداردهی کردیم.

برنامهٔ ۵: رمزگذاری برای ستون spending_score

```
score_mapping = {
    'Low': 0,
    'Average': 1,
    'High': 2
}
features['Spending_Score'] = features['Spending_Score'].map(score_mapping)
v features.head()
```

: (Nan) پر کردن داده های خالی

n_neighbors می خالی از KNNImputer در کتابخانه sikitlearn در کتابخانه استفاده کردیم به این صورت که KNNImputer داشته که یکی از پارامترهای این تابع است را ۸ انتخاب کردیم و این به این معناست که هر ردیفی که مفدار خالی داشته باشد میاید و ۸ ردیف نزدیک به ان را پیدا میکند و سپس از مقادیر ستون مورد نظر (یعنی ستونی که مقدار خالی در ان ردیفها میانگین گرفته و انرا به عنوان مقدار جدید سلول خالی قرار میدهد.

برنامهٔ ۶: پر کردن سلولهای خالی

```
from sklearn.impute import KNNImputer
imputer = KNNImputer(n_neighbors=8)
imputed_X = imputer.fit_transform(features)
features = pd.DataFrame(imputed_X, columns=features.columns)
```

١. پياده سازي الگوريتم ژنتيک:

الگوریتم ژنتیک یک روش جستجو و بهینهسازی مبتنی بر تکامل طبیعی است که در آن مفاهیمی از ژنتیک مانند انتخاب طبیعی، ترکیب و جهش برای یافتن بهترین جواب به کار گرفته می شوند. این الگوریتم در مسائل مختلف از جمله بهینهسازی، یادگیری ماشین و مسائل ترکیبیاتی مورد استفاده قرار می گیرد.

مراحل الگوريتم ژنتيك:

(آ) مقداردهی اولیه (Initialization):

مجموعهای از کروموزومها (یا راهحلهای ممکن) بهصورت تصادفی ایجاد میشود.

درس مبانی هوش محاسباتی صفحه ۹ از ۱۵

برنامهٔ ۷: مقداردهی اولیه

```
def initialize(count, column_count):
    zero_array = np.zeros((1, column_count), dtype=int)
    my_set = set()

while len(my_set) != count:
    random_numbers = random.sample(range(0, 21), column_count)
    copied_zero_array = zero_array.copy()
    for i in range(column_count):
        copied_zero_array[0][i] = random_numbers[i]
    my_set.add(tuple(copied_zero_array[0]))
    my_list = list(my_set)
    return my_set
```

(ب) ارزیابی برازندگی (Fitness Evaluation):

هر کروموزوم بر اساس یک تابع برازندگی ارزیابی میشود تا میزان تطابق آن با هدف مشخص شود. در کدی که ما زدیم تابع برازندگی، accuracy_score مدل است.

(ج) انتخاب (Selection)

کروموزومهای بهتر شانس بیشتری برای انتخاب شدن و انتقال به نسل بعدی دارند. روشهای مختلفی برای این کار وجود دارد، از جمله:

چرخ رولت (Roulette Wheel Selection) انتخاب بر اساس رتبه (Rank-Based Selection) انتخاب تورنمنت (Tournament Selection) که ما هر ۳تای انها را پیاده سازی کردیم.

برنامهٔ ۲: roulette-wheel-selection

درس مبانی هوش محاسباتی صفحه ۱۰ از ۱۵

برنامهٔ ۹: rank-based-selection

```
def rank_based_selection(initialize_gen, fitness_list, repeatable, count):
   fitness_list = get_rank_array(list(fitness_list))
   indices = list(range(len(fitness_list)))
   if repeatable:
       #repeatable
        selected_index = random.choices(indices, weights=fitness_list,
   k=count)
   else:
       #non-repeatable
        selected_index = np.random.choice(indices, size=count,
   replace=False, p=fitness_list/sum(fitness_list))
    selected_ones = list()
   initialize_gen = list(initialize_gen)
   for i in selected_index:
        selected_ones.append(initialize_gen[i])
   return selected_ones
```

برنامهٔ ۱۰ tournament-selection

```
def tournament_selection(initial_gen, gen_fitnesss, replacement, count,
   sample=2):
    selected_population = []
   if replacement:
        for _ in range(count):
            tournament_indices = random.choices(range(len(initial_gen)),
   k=sample)
            winner_index = max(tournament_indices, key=lambda i:
   gen_fitnesss[i])
            selected_population.append(initial_gen[winner_index])
        not_used = list(range(len(initial_gen)))
        thrown = []
        while len(not_used) >= sample and len(selected_population) < count:</pre>
            tournament_indices = random.sample(not_used, sample)
            winner_index = max(tournament_indices, key=lambda i:
   gen_fitnesss[i])
            selected_population.append(initial_gen[winner_index])
            for idx in tournament_indices:
                not_used.remove(idx)
                if idx != winner_index:
                    thrown.append(idx)
```

درس مبانی هوش محاسباتی صفحه ۱۱ از ۱۵

```
pool = thrown + not_used

while len(selected_population) < count and pool:
    if len(pool) >= sample:
        tournament_indices = random.sample(pool, sample)
    else:
        tournament_indices = random.choices(pool, k=sample)
    winner_index = max(tournament_indices, key=lambda i:
    gen_fitnesss[i])
    selected_population.append(initial_gen[winner_index])

return selected_population
```

(د) ترکیب (Crossover):

```
کروموزومهای انتخاب شده با یکدیگر ترکیب می شوند تا فرزندان جدید تولید شود. روشهای متداول شامل: ترکیب تکنقطهای (Single-Point Crossover) ترکیب دونقطهای (Two-Point Crossover) ترکیب دونقطهای (Uniform Crossover) که ما هر ۳تای اینها را هم پیاده سازی کردیم.
```

برنامهٔ ۱۱ : single-point-crossover

```
def single_point_crossover(parents):
    offspring = []
   for i in range(0, len(parents), 2):
        if i + 1 < len(parents):</pre>
            set1 = set()
            set2 = set()
            p1, p2 = parents[i], parents[i + 1]
            child_size = len(p1)
            cross_point = random.randint(1, len(p1) - 1)
            make1 = p1[:cross_point] + p2[cross_point:] + p1[cross_point:]
   + p2[:cross_point]
            make2 = p2[:cross_point] + p1[cross_point:] + p2[cross_point:]
   + p1[:cross_point]
            for j in make1:
                if len(set1) == child_size:
                    break
                set1.add(j)
            for j in make2:
                if len(set2) == child_size:
                    break
                set2.add(j)
            child1 = list(set1)
            child2 = list(set2)
```

```
offspring.extend([child1, child2])
return offspring
```

برنامهٔ ۲۲ : two-point-crossover

```
def two_point_crossover(parents):
    offspring = []
    for i in range(0, len(parents), 2):
        if i + 1 < len(parents):</pre>
            set1 = set()
            set2 = set()
            p1, p2 = parents[i], parents[i + 1]
            child_size = len(p1)
            point1 = random.randint(1, len(p1) - 2)
            point2 = random.randint(point1 + 1, len(p1) - 1)
            make1 = p1[:point1] + p2[point1:point2] + p1[point2:] +
   p2[:point1] + p1[point1:point2] + p2[point2:]
            make2 = p2[:point1] + p1[point1:point2] + p2[point2:] +
   p1[:point1] + p2[point1:point2] + p1[point2:]
            for j in make1:
                if len(set1) == child_size:
                    break
                set1.add(j)
            for j in make2:
                if len(set2) == child_size:
                    break
                set2.add(j)
            child1 = list(set1)
            child2 = list(set2)
            offspring.extend([child1, child2])
    return offspring
```

برنامهٔ ۱۳ : uniform-crossover

درس مبانی هوش محاسباتی صفحه ۱۳ از ۱۵

```
else:
                make1.append(g1)
                make2.append(g2)
        make1 += p1 + p2
        make2 += p1 + p2
        set1 = set()
        set2 = set()
        for gene in make1:
            if len(set1) == child_size:
                break
            set1.add(gene)
        for gene in make2:
            if len(set2) == child_size:
                break
            set2.add(gene)
        offspring.append(list(set1))
        offspring.append(list(set2))
return offspring
```

(ه) جهش (Mutation):

برخی از کروموزومها دچار تغییرات جزئی میشوند تا از همگرایی زودرس جلوگیری شود و تنوع حفظ گردد.

سرنامهٔ ۱۴: mutate

```
def mutate(individual, max_value):
   index = random.randint(0, len(individual) - 1)
   new_value = random.randint(1, max_value)
   while new_value in individual:
        new_value = random.randint(1, max_value)
   individual[index] = new_value
   return individual
```

(و) تكرار (Iteration):

همهی مراحل الف تا ه را دوباره انجام میدهیم تا به یکی از دو شرط زیر برسیم: رسیدن به حداکثر تعداد تکرار (iterations) رسیدن به نقطهی همگرایی(عدم بهبود برای مدت طولانی)

ارس مبانی هوش محاسباتی صفحه ۱۴ از ۱۵

حالا با استفاده از الگوریتمهایی که در بالا توضیح داده و پیاده سازی کردیم. الگوریتم ژنتیک را پیادهسازی میکنیم.

برنامهٔ ۱۵: Algorithm Genetic

```
def GA(X, y, count, column_count, repeatable, select_count,
   selection_type='roulette_wheel_selection', crossover='single_point',
   max_generations=100, target_accuracy=0.95, max_stagnation=50):
   initial_gen = initialize(count, column_count)
   generation = 0
   best_fitness = 0
   stagnation_counter = 0  # Track generations without improvement
   max_fitness_list = list()
   while generation < max_generations and best_fitness < target_accuracy:
        fitness_list = list()
       for i in range(len(initial_gen)):
            fitness_list.append(get_fitness(initial_gen, X, i, y))
        current_max = max(fitness_list)
        previous_best = best_fitness
        best_fitness = max(best_fitness, current_max)
       if best_fitness == previous_best:
            stagnation_counter += 1
        else:
            stagnation_counter = 0
        max_fitness_list.append(current_max)
        if stagnation_counter >= max_stagnation:
            break
        if selection_type == 'roulette_wheel_selection':
            selected_ones = roulette_wheel_selection(initial_gen,
   fitness_list, repeatable, select_count)
        elif selection_type == 'rank_based_selection':
            selected_ones = rank_based_selection(initial_gen,
   fitness_list, repeatable, select_count)
        elif selection_type == 'tournament_selection':
            selected_ones = tournament_selection(initial_gen,
   fitness_list, repeatable, select_count)
        else:
            raise ValueError('Invalid selection')
        if crossover == 'single_point':
            offspring = single_point_crossover(selected_ones)
```

درس مبانی هوش محاسباتی صفحه ۱۵ از ۱۵

```
elif crossover == 'two_point':
              offspring = two_point_crossover(selected_ones)
          elif crossover == 'uniform':
              offspring = uniform_crossover(selected_ones)
          else:
              raise ValueError('Invalid crossover')
          mutated_offspring = [mutate(list(child), column_count * 2) for
۴۵
     child in offspring]
          initial_gen = mutated_offspring
          generation += 1
      plt.plot(max_fitness_list)
      plt.xlabel('Generation')
۵۰
      plt.ylabel('Best Fitness')
      plt.title('Best Fitness per Generation')
      plt.show()
۵٣
۵۴
      return mutated_offspring
```