

پروژه ژنتیک

قسمت اول (

اول از همه دیتاست را بارگیری کرده و نمودار آن را بر اساس طولهای گلبرگ و کاسبرگ مشاهده می کنیم.

سپس توابع مختلف را تعریف می کنیم. روش کار بدین صورت است که ابتدا جمعیت اولیه ای از کروموزوم ها را تشکیل می دهیم. هر کروموزوم طولی به طول داده ها دارد (150) و به هر اندیس آن به صورت تصادفی عددی از 1 تا k اختصاص داده می شود (این که آن عضو از کدام گروه است) سپس به تعداد نسلیها تکامل را انجام داده و بهترین فرزند را ذخیره می کنیم.

معیار تابع برازش معکوس وزن دار واریانس نقاط خوشه هر داده است. برای هر خوشه یک centroid تعریف می شود که برابر با میانگین نقاط آن است و واریانس هم نسبت به این نقطه سنجیده می شود. باز ترکیب بدین صورت است که یک نقطه برش به صورت رندوم انتخاب می شود و برای فرزند اندیسهای قبل از آن نقطه از والد اول و بقیه از والد دوم به ارث می رسد. جهش نیز بدین با صورت است که احتمالی (اینجا 0.35) اتفاق می افتد؛ اندیسی از فرزند به طور تصادفی انتخاب شده و lable آن با لیبل نزدیکترین centroid عوض می شود. بدین ترتیب بعد از حدود 1000 نسل تکامل به جواب خوبی می رسیم. برای k های مختلف امتحان می کنیم و به نظر می رسد برای $k = 3$ پاسخ از همه بهتر است (k تعداد خوشه هاست)

در الگوریتم k_means در ابتدا به تعداد خوشه ها centroid تولید می شود و هر داده از کل داده ها بر اساس نزدیک ترین Centroid خوشه بندی می شود. سپس واریانس محاسبه می شود. در مرحله بعدی centroid های جدید میانگین داده های هر خوشه هستند و مرحله قبلی تکرار می شود. این کار آن قدر انجام می شود تا واریانس ها از حدی کمتر شوند (یا شرطی دیگر برای خاتمه) و بهترین نتیجه ذخیره می شود. این کار چند بار با centroid های اولیه مختلف انجام می شود تا به بهترین نتیجه برسیم.

این الگوریتم چند بار شروع مجدد دارد و چون جهش هم در آن وجود ندارد، احتمال دارد که در نقطه اکسترمم محلی گیر کند و به جواب مطلوبی نرسد.

قسمت دوم (

کلیات الگوریتم مثل قسمت اول است. جمعیت اولیه ای از کروموزوم ها را در نظر می گیریم. طول هر کروموزوم یکی از تعداد شهرها کمتر بوده و جایگشتی بین 2 تا تعداد شهرهاست. (به معنای اینکه از شهر یک شروع کرده و از این شهرها گذشته و سپس دوباره به شهر یک باز گردیم. برای سادگی یک ها حذف شده اند) تابع برازش برابر اندازه مسیر است. باز ترکیب بدین صورت اتفاق می افتد که ابتدا نقطه ای برای برش در نظر گرفته می شود. سپس از آن اندیس به بعد از والد اول و اندیس های باقیمانده با حفظ ترتیب از والد دوم به ارث می رسد. (برای مثال اگر والدین 2354 و 3425 باشند، فرزند اول 3254 (54 از والد اول و 2 و 3 با حفظ ترتیب از والد دوم به ارث می رسند) فرزند دوم نیز 3425 خواهد بود). جهش هم بدین گونه است که با احتمالی جای دو اندیس با یکدیگر عوض می شود.

در نهایت به صورت رندوم شهر ها و فاصله بین آنها را مقاردهی کرده و الگوریتم را اجرا می کنیم و بهترین نتیجه را چاپ می کنیم.