دکتر: "کمک"

توی این خواندنی می خوایم یکم به دکتر ها برای تشخیص سرطان پوستی که اسمش ملانوما هست کمک کنیم.

حالا یه سوال دکترا چطوری تشخیص می دن؟ مگه تشخیص دکتری چه مشکلی داره که از کامپیوتر بخوایم تشخیص بده یا کمکمون کنه که تشخیص بدیم؟ 😊

اول بیاین به مشکل بزرگ که از گذشته های دور بوده اشاره کنیم ( ترس از دکتر ☹ ) چرا از دکتر می ترسیم چون دکتر ها درد دارن در واقع آزمایش ها و عمل ها و .... درد دارن و یه مشکل بزرگ تر اینه که اگه سالم هم باشی باید درد آزمایش ها رو تحمل کنی

ولی دیدین که عکس گرفتن چه حالی میده ( در واقع هیچی حس نمی کنید ولی کیف میده 😊 )

حالا ما می خوایم چی کار کنیم؟؟ ما می خوایم یکی از اون آزمایش ها رو به راحتی عکس گرفتن بکنیم .

حالا که فهمیدیم کار ما عجب کمک بزرگی به بشریت می کنه بریم سراغ اینکه اصلا این بیماری چی میگه

خوب سرطانی که ما می خوایم بررسی کنیم با چشم هم قابل دیدنه 😊 و شبیه یه لکه روی پوست هست (ولی هر لکه ای غده سرطانی نیست)

حالا ما می خوایم به آقا / خانوم دکتر بگیم که از اون لکه ای که روی پوست مریضشون هست یه عکس بگیرن به برنامه ی ما بدن تا ما یه کمکی بهشون بکنیم که بفهمن این یه غده ی سرطانیه ☹ یا سرطانی نیست😊

حالا ما چه کمکی می خوایم بکنیم و چه جوابی به عنوان جواب آزمایش به دکتر می خوایم بدیم؟؟

خوب ما اول میایم و لکه رو توی عکس پیدا می کنیم حالا محیط و مساحتشو به واحد پیکسل محاسبه می کنیم بعد میریم یه دایره می کشیم که مساحتش با مساحت لکه ی ما یکی باشه بعد نسبت محیط لکه رو به محیط دایره ی هم مساحتش محاسبه می کنیم و اون مقدار رو ration صدا می زنیم( اسمشه )

حالا هم مساحت و هم محیط و هم ratio رو به دکتر اعلام می کنیم

به همین سادگی 😊

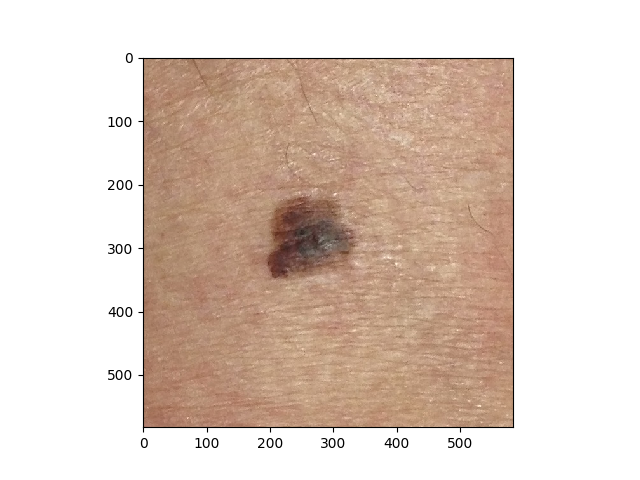
ولی حالا این دکتر از این مقدار چی می فهمه ؟؟

خوب یه سوال دیگه اگه این مقدار زیاد باشه معنیش چیه ؟؟ معنیش اینه که این لکه محیطش نسبت به مساحتش خیلی زیاده و حالا این معنیش چیه یعنی تو رفتگی ها و ناهمواری های زیادی روی محیط لکه وجود داره 😐 ( خوب بقیشم برین از دکترها بپرسید )

حالا ما باید در مورد داده هامون که همون عکس ها هستند حرف بزنیم .

ما یه عالمه عکس از لکه های پوستی داریم که دکتر ها با روش های وحشتناک دردناک خودشون اونها رو بررسی کردن و به ما گفتن که آیا سرطان دارن یا امکانش وجود داره سرطانی باشن و یا امکانش خیلی کمه که سرطانی باشن( اره دیگه دکتر ها خودشون هم از روش های ترسناک خودشون خسته شدن و می خوان به ما کمک کنن تا دیگه مردم ازشون نترسن . اونها هم دل دارن )

این اون عکسیه که می خوایم توی این خواندنی بررسی کنیم



حالا بریم یکم درمورد عکس توضیح بدیم( منظورم از عکس هر عکسیه حتی عکس شب تولد شما دوست عزیز)

یه عکس از نقاط خیلی خیلی کوچولویی تشکیل شد که بهش می گن پیکسل . هر کدوم از این نقاط فقط یک رنگ می تونه داشته باشه مثلا می تونه فقط آبی باشه

برای اینکه بهتر بفهمین اگه تو خونه هستین زیر پاتون رو نگاه کنید چی می بینید ( اسمش قالیه) اگه یکم ریز بشین می بینید که یه عالمه نخ هستند که ما فقط سرشون رو می بینیم و هر نخ یه رنگ هست و وقتی از دور به قالی نگاه می کنیم یه تصویر رنگارنگ می بینیم اون نخ ها همون پیکسل های قالین عکس هم دقیقاً از یه همچین چیزایی تشکیل شده.

خوب حالا که با پیکسل آشنا شدیم بریم سراغ رنگ ها

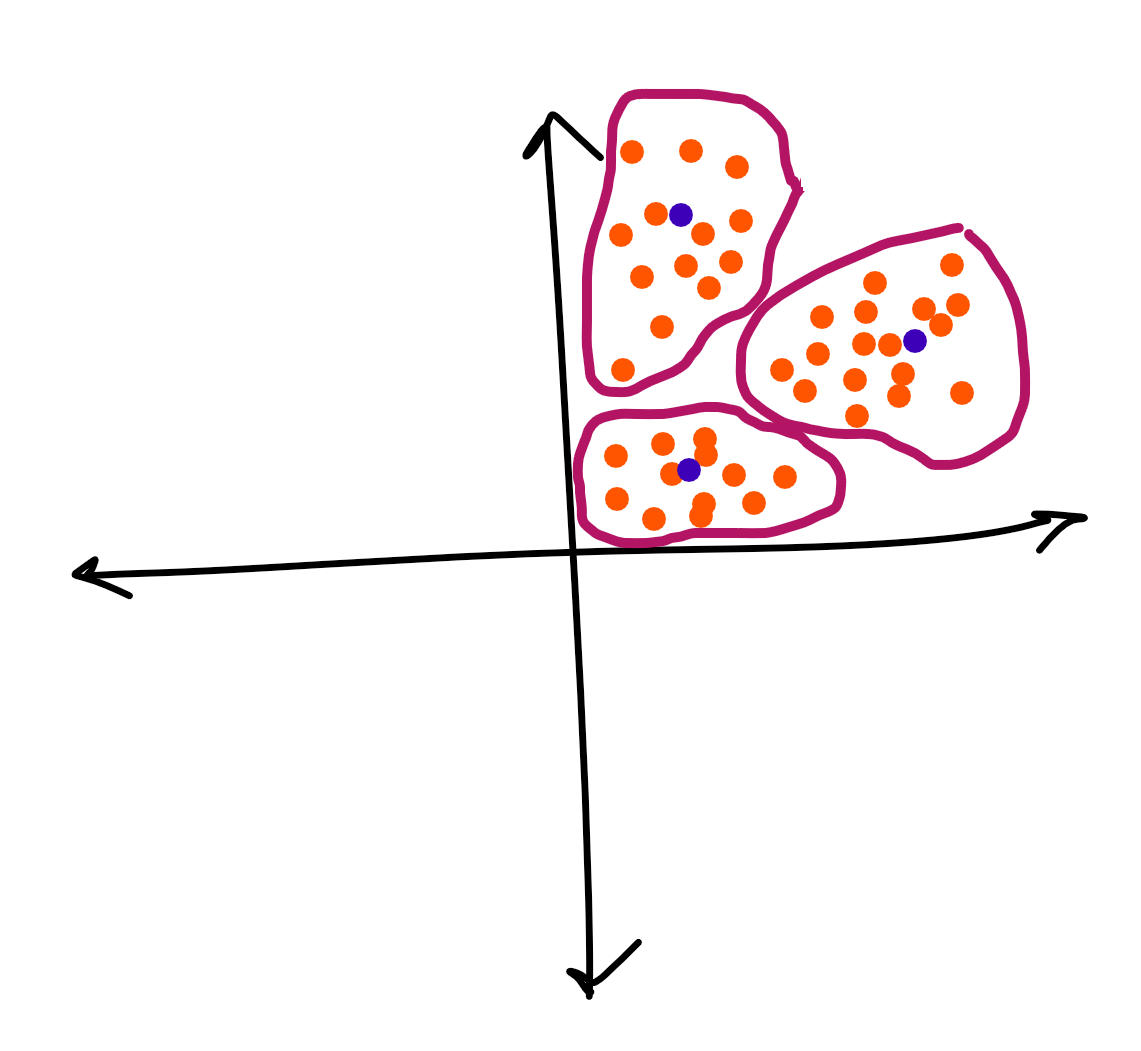
می دونیم که هر رنگ از مجموع سه رنگ ابی و قرمز و سبز تشکیل شده (RGB) در واقع هر رنگ رو میشه به این سه تا رنگ تجزیه کرد .

پیکسل سه مقدار داره که هر کدوم نشون دهنده ی مقدار هر کدوم از این سه رنگ در رنگ اون پیکسله .

هر مقدار توی هر پیکسل یک عدد صحیح از 0 تا 255 هست( یعنی 256 تا ) و چون سه تا از این مقدار ها توی هر پیکسل وجود داره 256 \* 256\*256 حالت ( رنگ) مختلف یه پیکسل می تونه به خودش بگیره یعنی چند تا ؟؟ یعنی خیلی زیاد و این خودش کار پردازش رو خیلی سخت می کنه و تقریبا امکان نداره توی یه عکس از شب تولد شما همه بخش های لباس شما توی عکس یه رنگ داشته باشن یعنی پیکسل هایی که دارن عکس شما رو نمایش میدن همه یه رنگ داشته باشن (هر چند که کل لباس شما یه رنگ داره ) خوب اگه گفتین چرا اینطوری میشه؟؟

برای همین اولین کاری که ما می خوایم بکنیم اینه که دامنه ( تعداد حالات ممکن) رنگ های موجود توی عکس رو کاهش بدیم ولی به چند تا رنگ ؟؟ من با توجه به اینکه عکس هام فقط بخشی از پوست و کل لکه رو نشون می دن به 3 تا رنگ کاهش می دم ( چرا 3 تا ؟؟) خوب یکی رنگ پوست یکی رنگ لکه و یکی هر چیز دیگه ای که مثل نویز توی عکس من هست. (ولی شما دوست داری تغییرش بده)

خوب برای این کار من از الگوریتم KMeans استفاده کردم ( یا خدا این چیه دیگه 😐) چیز خاصی نیست

یه روش برای تقسیم کردنه فرض کنید یه عالمه نقطه توی یک صفحه مختصات دارین می خواین این نقطه ها رو به 3 دسته تقسیم کنیم یعنی چی یعنی اونهایی که نزدیک به هم هستن رو توی یک دسته قرار بدیم و یه سرگروه که وسط همه ی اونهاست رو انتخاب کنیم از این به بعد اگه ازمون پرسیدن فلان نقطه مختصاتش چیه ما مختصات سرگروهشو رو اعلام می کنیم 

خوب توی این نمودار بالا نگاه کنید اون اّبی ها سرگروهن دسته ها هم مشخص شده حالا اگه یکی از ما مختصات هر کدوم از اون نقاط رو بپرسه ما مختصات سرگروه رو میگیم این طوری ما سه جواب بیشتر برای گفتن نداریم و این خودش خیلی خوبه 😊

خوب حالا برای رنگ ها هم همین کار رو می کنیم. یعنی چی؟ میایم هر پیکسل رو یه نقطه در نظر می گیریم و هر کدوم از اون سه رنگ رو یکی از بٌعد هاش قرار می دیم یعنی رنگ رو توی یه صفحه ی مختصات سه بعدی نمایش می دیم ( اینکه 3 بعدی شد ) اشکالی نداره هم x داره هم y داره هم z که ب همون مقدار R, G, B پیکسل ما هستند خوب ما میایم اینها رو خوشه بندی می کنیم( همون گروه گروه کردنه ) و به جای رنگ هر کدوم از پیکسل ها رنگ سر گروهشو قرار می دیم این طوری عکس ما فقط سه نوع رنگ خواهد داشت

حالا اینکه الگوریتم ما چطوری این کار رو انجام میده رو شما باید برین تحقق کنید. ( خیلی سخت نیست کافیه یه سرچ انگلیسی کوچیک بکنید)

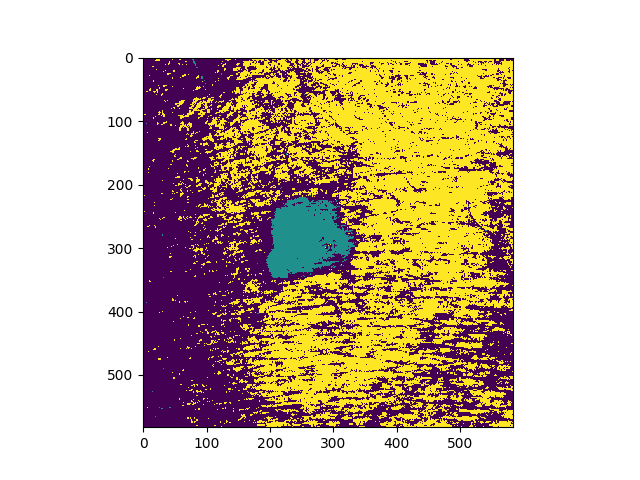
ولی من توی یه فایل پایتون برنامه ای نوشتم که این کارو بکنه اسم فایل هم Clustering هست

توی این فایل یه کلاس هست که دوتا ورودی داره یکی image که همون عکسه و یکی هم n\_clusters که تعداد خوشه هاست که به صورت پیش فرض مقدارش 3 هست

آخرشم باید متد run رو اجرا کنید و خروجی این متد هم اولیش عکسه دومی هم مرکز خوشه هاست.

اگر هم می خواین خود عکس که تبدیل شده رو به راحتی ببینید متد show\_image رو از این کلاس اجرا کنید.

ولی حالا خروجیشو برای عکس ما ببینید

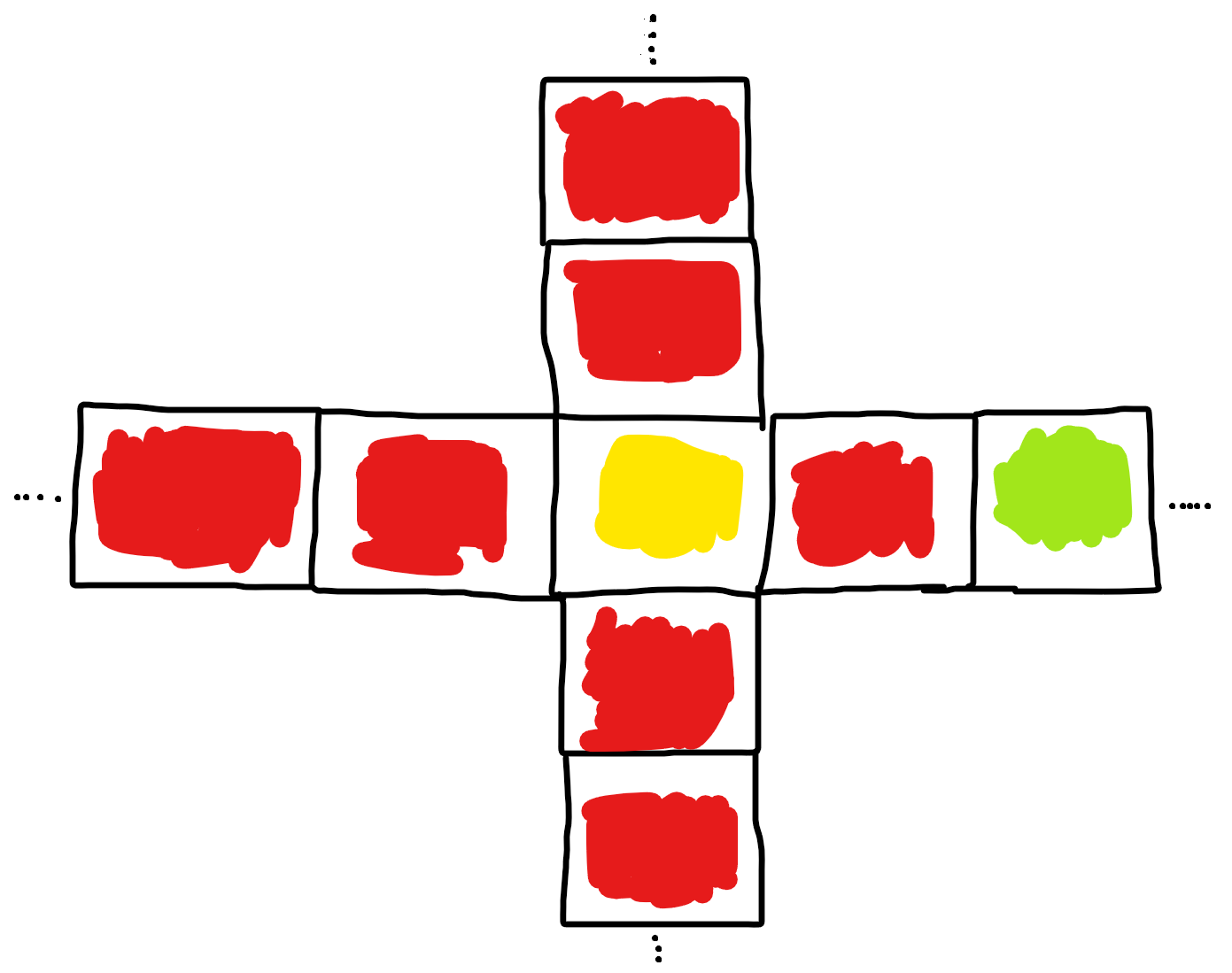


اینجا برای نمایش از 3 تا رنگ فرضی استفاده شده (برای اینکه بهتر نمایش بدیم) ولی در واقعیت هر پیکسل به رنگ سرگروهشه. 😊

خوب کار بعدی چیه؟؟

وای خدای من چرا همچینی شد یه جاهایی وسط لکه عضو لکه نیست و سرگروهش رنگ پوسته حالا چی کار کنیم ؟؟

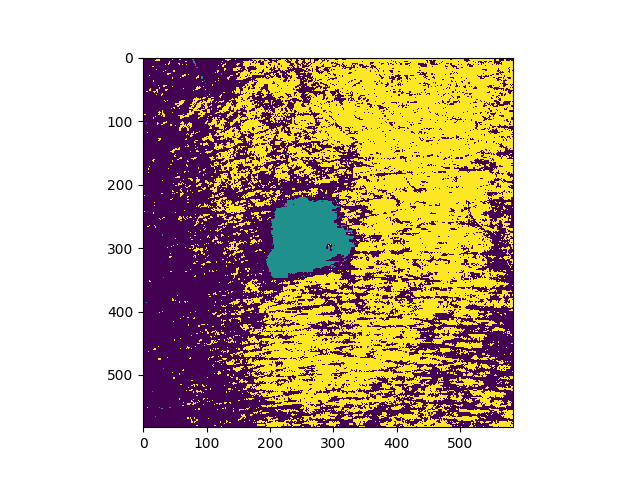
کار خاصی نمی کنیم میام برای هر پیکسل همسایه هاشو بررسی می کنیم که عضو کدوم گروهن اگه 70 درصدشون عضو یه گروه دیگه بودن پیکسل ما هم عضو اون گروه میشه (حالا چرا 70 درصد ؟؟ چون دوست داشتم شما دوست داری تغییرش بده دعوا نداریم که 😊 )



خوب اون بالا رو نگاه کنید اون زرده رو ما می خوایم بررسی کنیم ( زرد رنگ سر گروهشه) همسایه هاشو نگاه می کنیم می بینیم بیشترشون قرمزن به این نتیجه می رسیم که این یکی هم قرمزه (ولی حالا چرا تا 2 تا همسایه اونور تر ؟؟ خوب می نمی تونستم توی شکل بیشتر از اینو جا بدم ولی توی برنامه تا 4 تا اونور ترو بررسی کردم شما دوست داری بکنش 100 تا به من چه!!)خوب این طوری شد که کلی از پیکسل های فراری از گروهشون رو به گروه خودشون برگردوندیم و به یه لکه تقریبا خوب رسیدیم

اسم فایل پایتونی که این کارا رو انجام بده smooth هست. که یه کلاس به همین اسم داره که ورودی اولش همون عکسیه که خروجی فایل پایتون قبلی هست دومی هم سر گروه ها هستند سومی هم تعداد همسایه هایی که می خواین بررسی بشه ( همون m ) و آخری همون درصد تفاوت همسایه هاست ( اون که می دوست داشتم 70 در صد باشه) و اسم ورودی هم n هست مثلا 70 در صد رو باید بهش 0.7 بدیم

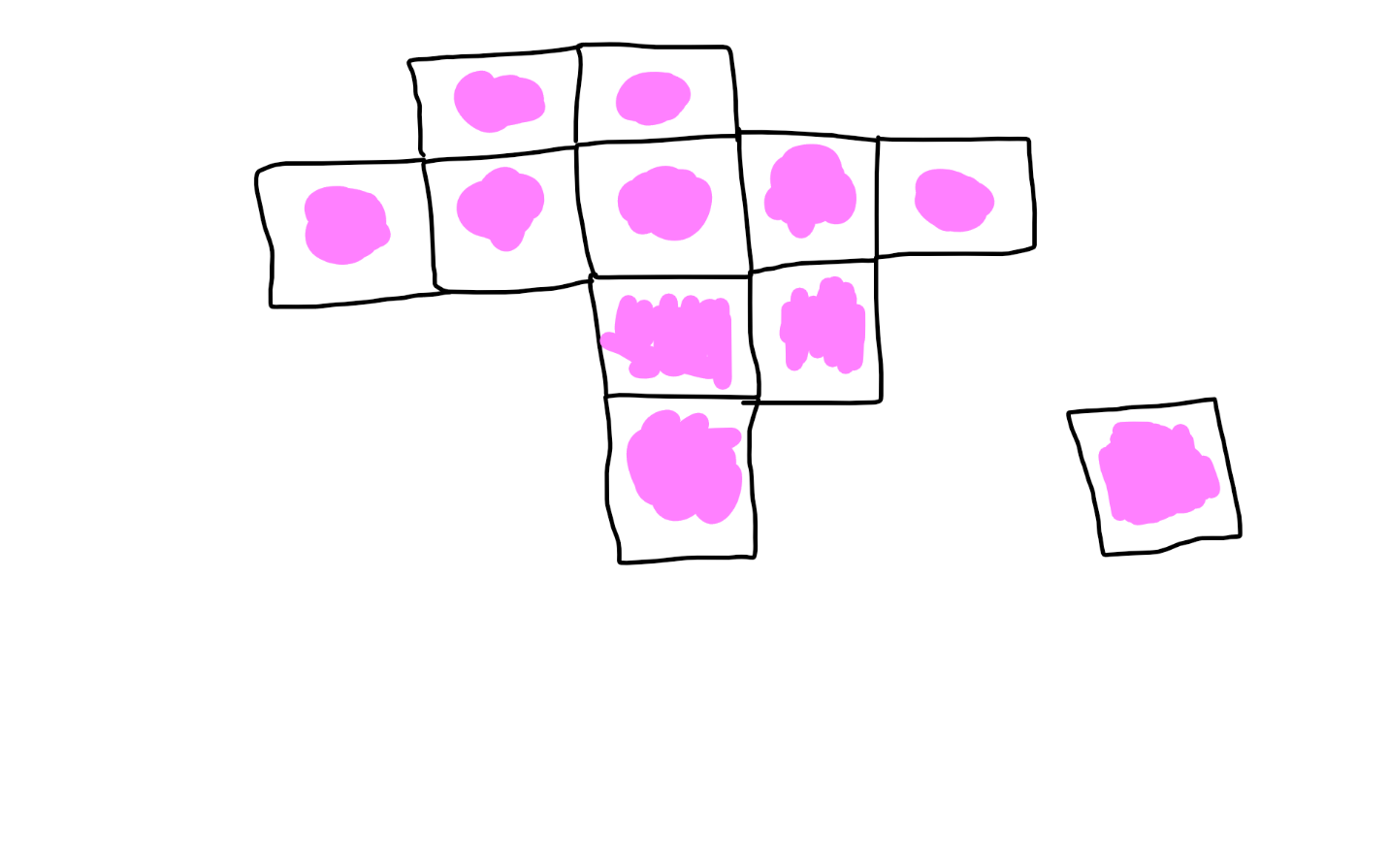
اینجا هم مثل کلاس قبلی یه متد run و یه متد show\_image داریم که خروجی run فقط عکس جدید هست.



ولی حالا یه سوال ما 3 تا گروه داریم کدوم گروه لکه هست؟؟ من که می گم اون که از همه تیره تره حالا کدوم از همه تیر تره ؟؟ 😐 خوب باید بگم مقدار RGB توی رنگ سیاه [0,0,0] هست یعنی هیچ رنگی وجود نداشته باشه پس هر گروهی که سرگروهش به این مقدار ها نزدیک تر باشه اون گروه تیره تره ( درواقع میانگین مربعاتشون به هم نزدیک باشه) و اون گروهی که تیره تر باشه گروه لکه هست 😊

خوب لکه رو پیدا کردیم ولی ای دل غافل توی اون فیلتری که هر فراری رو به گروه خودش برگدوندیم همه ی فراری ها به خوشه ی( گروه) خودشون برنگشتن و هنوز هم فراری هایی وجود داره که دیگه باید بیخیالشون بشیم و بریم . میگین نه عکس بالایی رو نگاه کنید بالا سمت چپ چند تا پیکسل هستند که عضو گروه لکه ان ولی هیچ ربطی به لکه ندارن و با فاصله ی زیادی از اون قرار گرفتن ☹

ولی همین طوری هم رهاشون نمی کنیم میریم توی خوشه ی لکه و اونهایی رو که به بقیه چسبیده نیستن رو حذف می کنیم

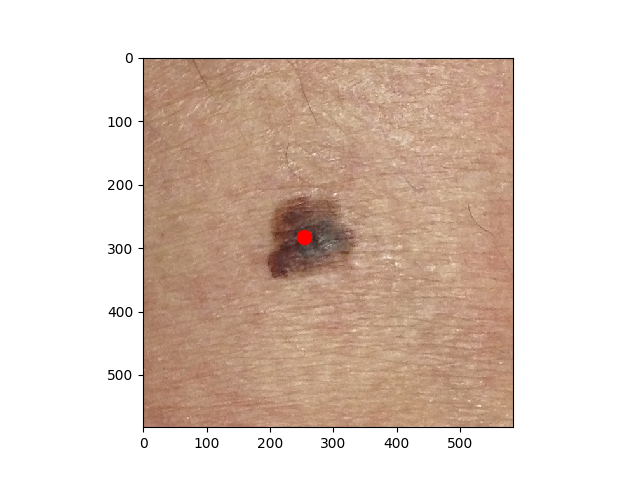


مثلا توی عکس بالا مرکز خوشه ( همون سرگروه) برای همه بنفشه ولی مثل اینکه یکی از اینها به بقیه چسبیده نیست پس ما اونو از خوشه حذف می کنیم

ولی خوب یه سوال !

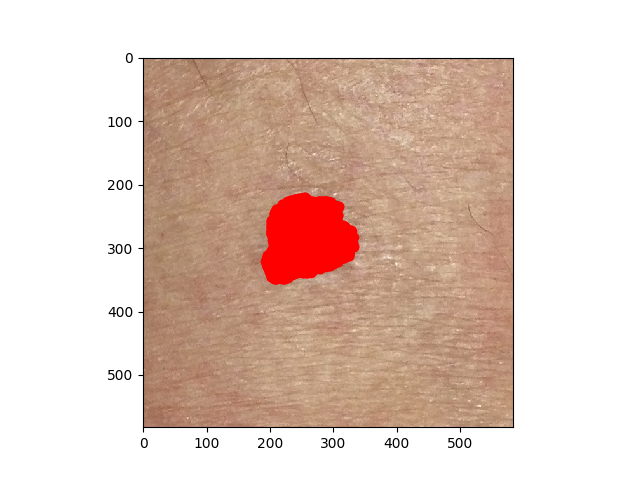
امکان داره دو دسته از پیکسل ها باشن که هر کدوم از چند تا پیکسل به هم چسبیده شده درست شده باشن بازم مشکل ☹ خوب ما برای اینکه بفهمیم کدوم یکی لکه هست و کدوم لکه نیست میایم و مرکز تجمع رو پیدا می کنیم یعنی چی؟ یعنی اونجایی که بیشتر پیکسل های لکه جمع شدن (مثل جا هایی که یه چیزی رو مفتی می دن و یهویی شلوغ میشه و تجمع میشه )

خوب من از طریق یه الگوریتم نه چندان سخت اون نقطه تجمع رو پیدا کردم .(البته منظورم میانگین نیست 😊 ولی از میانگین توی الگوریتمم استفاده کردم ) یکم گفتنش طولانیه اگه دوست داشتین می تونید برین و برنامه رو یه بار بخونید



حالا از اون مرکز شروع می کنیم و تمام همسایه هایی که بهش چسبیدن رو پیدا می کنیم .

واینطوری میشه:



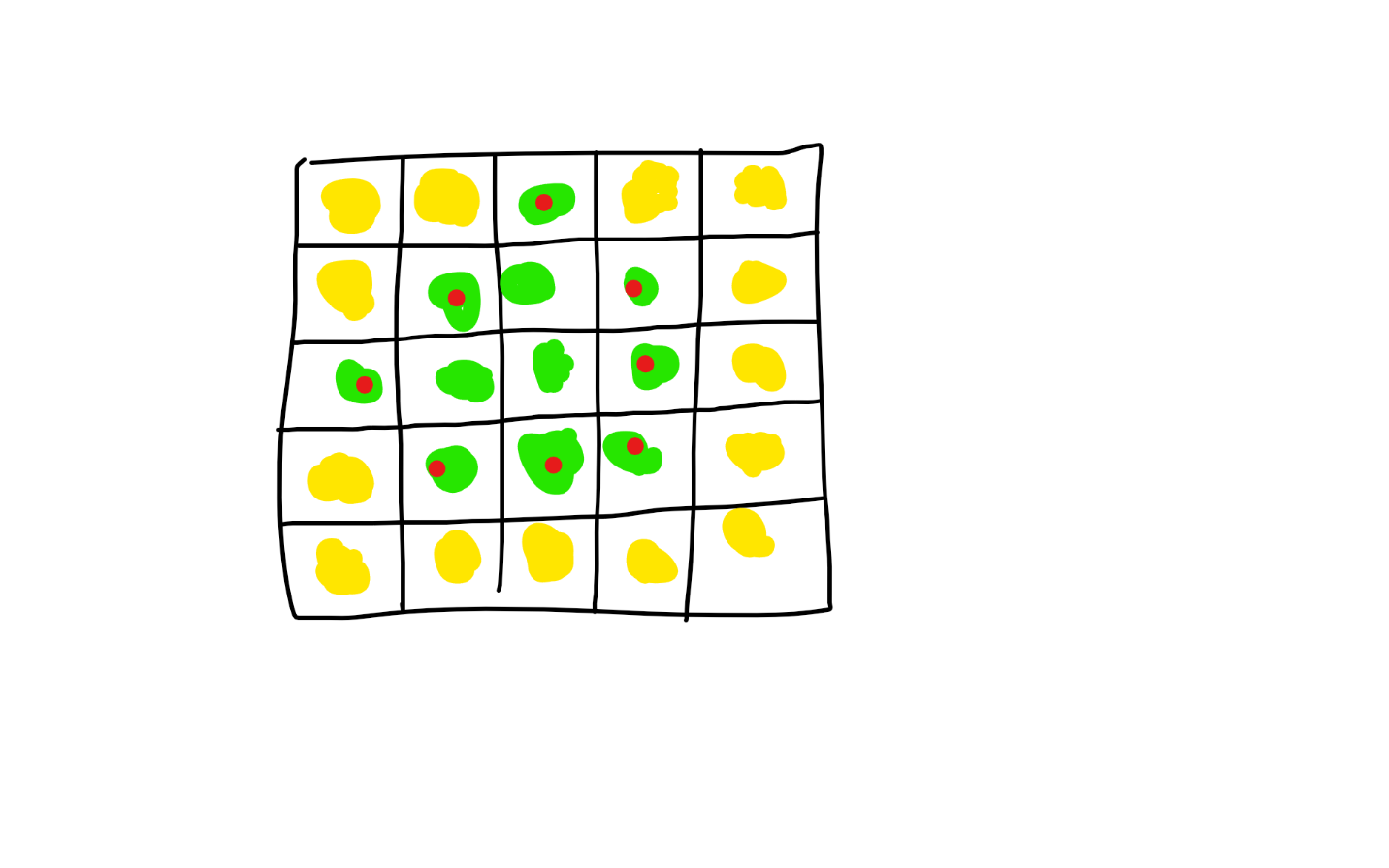
و این گونه می شود که ما لکه رو پیدا کردیم و می تونیم مساحتشون رو هم حساب کنیم

من این دوتا کار رو توی دوتا فایل انجام دادم یکی find\_median و یکی هم find\_mole که از اسم هر کدوم پیداست که چی کار می کنند و مثل کلاس های قبلی هم run دارن و هم show\_image و این خودش یه قانونه( البته توی این برنامه)

ولی پس محیط چی😐

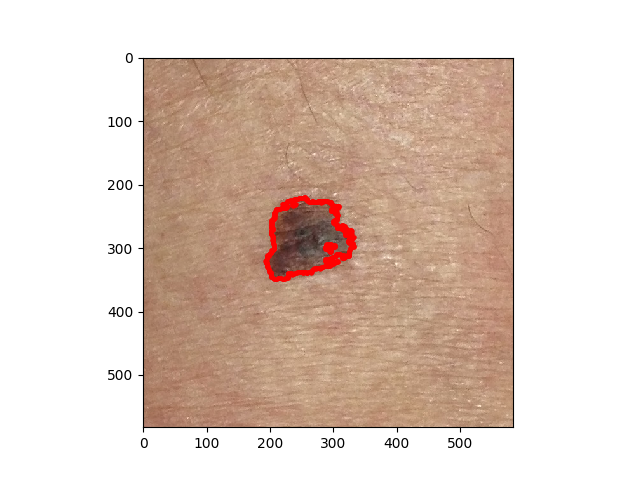
خوب بریم که محیط رو محاسبه کنیم حالا کدوم پیکسل ها پیکسل های محیط هستند ؟؟

خوب معلومه دیگه اون پیکسل هایی از لکه که یکی از همسایه های دیوار به دیوارش برچسبش ( همون سرگروهش) با برچسب لکه تفاوت داشته باشه .



مثلا توی عکس بالا سبز ها لکه ان و اونهایی که نقطه قرمز دارن محیط هستند

حالا بیاین این کارو برای عکسمون انجام بدیم

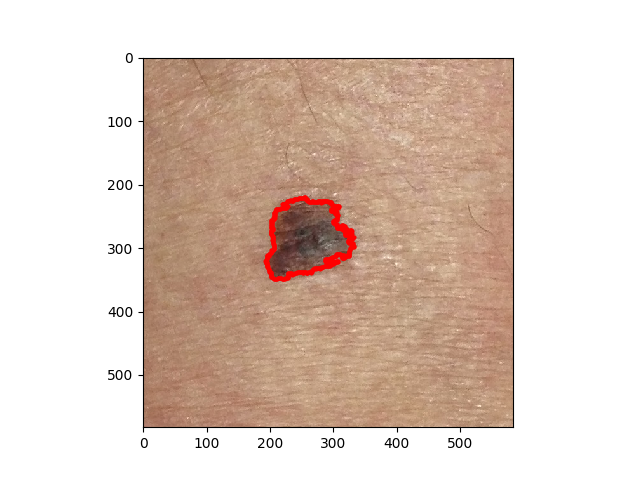


وای دیدین چی شد !!! یکی از اون نویز های نامرد باعث شد یه بخش هایی که محیط نیستند رو هم محیط حساب کنیم

حالا چی کار کنیم باید دنبال خط های دنباله دار بگردیم مثل چی؟؟ فرض کنید ما یه دسته آدم دارین که یه سری ها توی صف وایستادن یه سری آدم دیگه هم با کمی فاصله از اون صف وایستادن یا حتی با پر رویی تمام یه صف کوچیک دیگه تشکیل دادن تا هر موقع که شد بپرن وسط صف اصلیه ( همون بزنن تو صف خودمون) حالا ما می خوایم اون صف اصلی رو پیدا کنیم. چی کار می کنیم؟ میایم دنبال خط می گردیم و اگه دوتا خط دنباله دار شد اونی رو دنبال می کنیم که بزرگتره

و این گونه می شود که ما محیط واقعی را پیدا می کنیم ( البته اینجا الگوریتم من برای بعضی عکس ها خوب جواب نمی ده و همون فیلتر نشدش بهتر کار میده)

توی فایل perimeter محیط رو پیدا کردم و توی فایل filter\_perimeter دنبال خط های دنباله دار گشتم.



حالا از طریق اون روشی که اول خواندنی گفتم میایم و به دکتر نتیجه رو اعلام می کنیم .

و برای محاسبه هم از فایل compute\_ratio استفاده می کنیم

خوب حالا شاید به خواین یکم بیشتر درمورد برنامه نویسی این برنامه بدونید

توی این برنامه که به زبان پایتون نوشته شده چند تا کتابخونه استفاده شده که توی پایتون نیست و باید اونها رو نصب کنی که عبارتند از :

Numpy

Scipy

Matplotlib

SKLearn

پس برای اجرا برنامه به این چند تا کتابخونه نیاز دارین

حالا برنامه کجاست ؟

برنامه توی github هست که می تونی به راحتی اون رو دانلود کنید از کجا؟ از این لینک پایینی

<https://github.com/MMReisi/Melanoma>

البته پیشنهاد می کنم از Fork استفاده کنید چون احتمالا برنامه تغییرات خوبی توش ایجاد خواهد شد

خوب پروژه رو که دانلود می کنید اسمش Melanoma هست فولدرو باز می کنید یه فولدر دیگه توش هست اسمش Samples\_images هست و توش کلی عکسه که دکترا به ما دادن ( دستشون درد نکنه 😊)

یه فولدر دیگه هم به اسمش question که توش یه فایل هست که من با کمک از اون برنامه رو نوشتم این فایل مال دانشگاه پلی تکنیک توی ایتالیاست.

توی فولدر یه فایل دیگه هم گذاشتم که اسمش runner.py هست که شما باید اجراش بکنید وقتی اجاراش کردین اسم یه عکس می خواد (از همون عکس هایی که توی فولدر samples\_images هست ) فقط فراموش نکنید که فرمت رو هم بنویسید مثلا بنویسید low\_risk\_1.jpg برنامه ی من یکم کنده پس یکم باید صبر باشید سپس برنامه خروجی می ده و عکس ها رو به ترتیب نشون می ده برای رفتن از هر عکس به عکس بعدی باید عکسی که باز هست رو ببندید .

امیدوارم خوشتون اومده باشه 😊