به نام خدا

عنوان:

تکلیف دهم درس یادگیری ماشین

استاد:

دکتر پدرام

دانشجو:

محمدعلی مجتهدسلیمانی

شماره دانشجویی:

٤٠٣٣٩٠٤٥٠٤

تاریخ:

١٤٠٣/٠٩/٢٩

سوال ۱

سوال ①

الف) ✓ درست است

$$P(a|b,c) = \frac{P(a,b,c)}{P(b,c)}$$

$$P(b|a,c) = \frac{P(a,b,c)}{P(a,c)} = P(a,c)$$

$$P(b,c) = P(a,c) \longrightarrow P(b|c) = \frac{P(b,c)}{P(c)} = \frac{P(a,c)}{P(c)} = P(a|c)$$

تونستیم از روی یکی دیگری بدست بیاوریم.

ب) ✗ لازم نیست. عبارت می‌گوید که a مستقل است نسبت به c و b اما از روی این نمی‌توان نتیجه گرفت که b هم از c مستقل باشد. مثال

A = آمدن رو در سکه اول

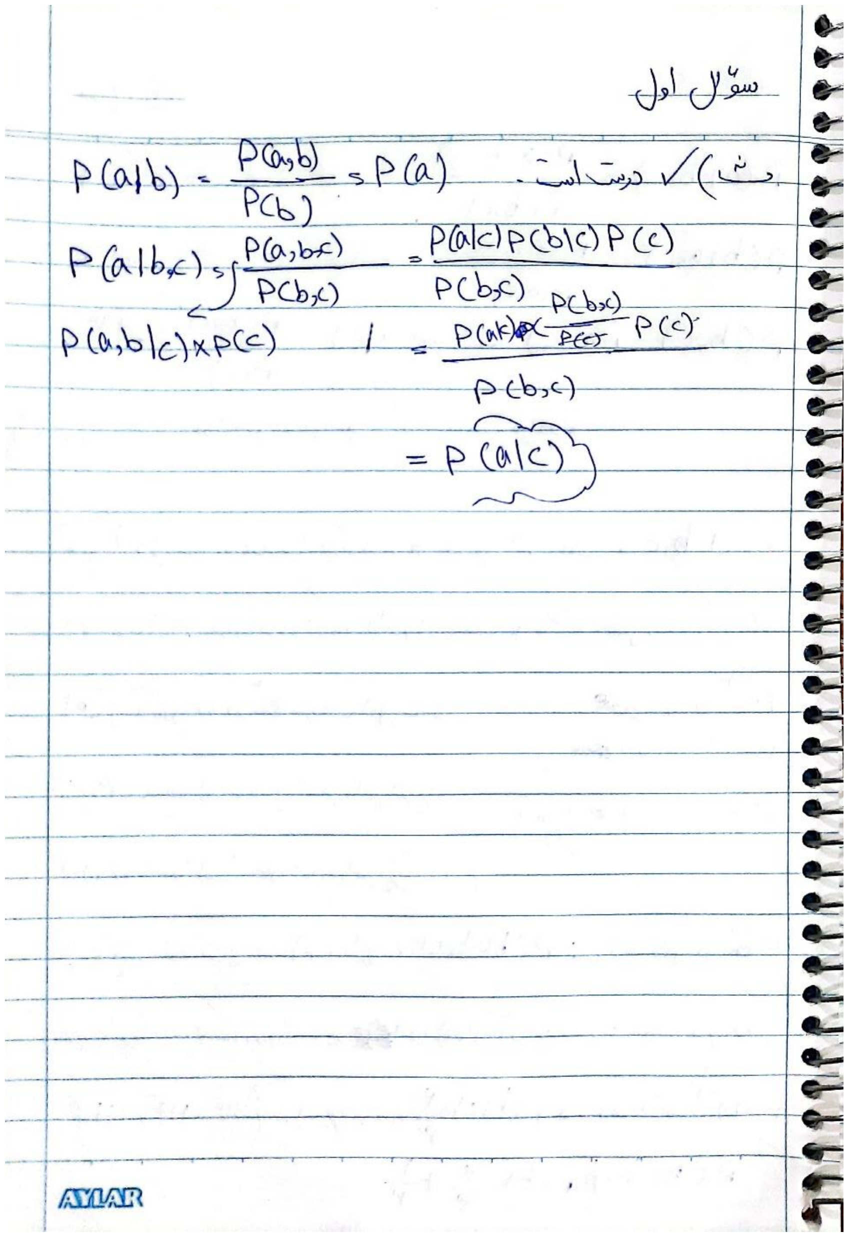
B = آمدن رو در سکه دوم

C = آمدن یک رو حداقل در دو پرتاب سکه

احتمال آمدن رو در پرتاب اول مستقل از B و C خواهد بود زیرا می‌دانیم در پرتاب اول رو آمده است حالا اینکه در پرتاب دوم یا C هست تأثیری در A ندارد. اما در $P(b|c)$ که می‌گوید فرض کن حداقل یک رو در پرتاب ۲ سکه به دست بیاید و براساس آن سکه دوم هم رو بیاید می‌شود $\{HH, HT, TH\}$ برابر می‌شود با $P(b|c) = \frac{2}{3}$ که متفاوت است با $P(b) = \frac{1}{2}$

$$P(b|c) \neq P(b) \quad \frac{2}{3} \neq \frac{1}{2}$$

AYLAR



سؤال اول

ب) ✓ درست است. $P(a|b) = \frac{P(a,b)}{P(b)} = P(a)$

$P(a|b,c) = \frac{P(a,b,c)}{P(b,c)} = \frac{P(a|c)P(b|c)P(c)}{P(b,c)} = \frac{P(a|c) \times \frac{P(b,c)}{P(c)} P(c)}{P(b,c)} = P(a|c)$

$P(a,b|c) \times P(c)$

همانطور که مشخص است مورد اول و سوم صحیح هستند و میتواند به اثبات برسد.

برای مورد دوم نمیتوانیم استقلال خواسته شده را اثبات کنیم و یک مثال نقض ارائه کرده ایم شرح مثال به این صورت است:

فرض کنیم پیشامد A رو آمدن در پرتاب اول سکه باشد.

پیشامد B رو آمدن در پرتاب دوم سکه باشد.

پیشامد C حداقل یکبار رو آمدن در پرتاب ۲ سکه باشد.

همانطور که مشخص است پیشامد A مستقل از B,C خواهد بود و بدون دانستن این احتمال شرطی هم میتوانستیم A را بدست بیاوریم. اما B مستقل از C نخواهد بود و مقادیر این تساوی با همدیگر مساوی نیست زیرا برای سمت چپ تساوی ما $2/3$ داریم چون ممکن به شرح زیر باشد: {رو رو، رو پشت، پشت رو} باشد که متفاوت از پیشامد B که احتمال آمدن رو $1/2$ است.

شرح کامل توضیحات در عکس آمده است.

سوال ۲

سؤال ۲

Stolen = Yes $\longrightarrow P(Y \mid Red, SUV, Domestic) = ?$

Stolen = No $\longrightarrow P(N \mid Red, SUV, Domestic) = ?$

$$P(Red \mid Y) = \frac{\frac{3}{10}}{\frac{5}{10}} = \boxed{\frac{3}{5}} \quad P(SUV \mid Y) = \frac{\frac{1}{10}}{\frac{5}{10}} = \boxed{\frac{1}{5}}$$

$$P(Domestic \mid Y) = \frac{\frac{2}{10}}{\frac{5}{10}} = \boxed{\frac{2}{5}}$$

$$P(Y \mid \underbrace{Red, SUV, Domestic}_{E}) = P(E \mid Y) \times P(Y) = \frac{3}{5} \times \frac{1}{5} \times \frac{2}{5} \times \frac{5}{10} = \boxed{\frac{3}{125}}$$

$$P(Red \mid N) = \frac{\frac{2}{10}}{\frac{5}{10}} = \boxed{\frac{2}{5}} \quad P(SUV \mid N) = \frac{\frac{3}{10}}{\frac{5}{10}} = \boxed{\frac{3}{5}}$$

$$P(Domestic \mid N) = \frac{\frac{3}{10}}{\frac{5}{10}} = \boxed{\frac{3}{5}}$$

$$P(N \mid \underbrace{Red, SUV, Domestic}_{E}) = P(E \mid N) \times P(N) = \frac{2}{5} \times \frac{3}{5} \times \frac{3}{5} \times \frac{5}{10} = \frac{9}{125}$$

stolen Yes $\frac{3}{125}$ < stolen No $\frac{9}{125}$

نمونه دزدیده نخواهد شد.

برای سوال ۲ برای نمونه جدید قاعده بیز را برای هر ۲ کلاس دزدیده شدن و دزدیده نشدن را محاسبه کردیم. ابتدا احتمال شرطی را برای هر کلاس بدست آوردیم و بعد با استفاده از داشتن دانش پیشین (prior) و محاسبه likelihood توانستیم قاعده بیز را حساب بکنیم و به این نتیجه رسیدیم از آنجایی که احتمال کلاس دزدیده نشدن بیشتر است پس این نمونه دزدیده نخواهد شد و برچسب این کلاس NO خواهد بود.

سوال ۳

سوال ۳

الف)

team A wins 65% = P(A)

team B wins ۳۵% = P(B)

team A won ۳۵% were on team B'field = P(B'field | A) = ۳۵%

team B won ۷۵% matches at their home = P(B'sfield | B) = ۷۵%

P(B | B'sfield) = ?

Bayes Rule

$$P(B \mid B'sfield) = \frac{P(B'sfield \mid B)\, P(B)}{P(B'sfield)}$$

$$P(B'sfield) = P(B'sfield \mid A) \times P(A) + P(B'sfield \mid B) \times P(B)$$

$$P(B'sfield) = (۰٫۳۵ \times ۰٫۶۵) + (۰٫۷۵ \times ۰٫۳۵) = ۰٫۴۹$$

$$P(B'sfield \mid B) \times P(B) = \boxed{۰٫۲۶۲۵}$$

ب)

$P(B|B's\ field) =$ ۰٫۲۶۲۵

$P(A|B's\ field) = ?$

$P(A|B's\ field) = P(B's\ field|A) \times P(A) =$ (۰٫۳۵ × ۰٫۶۵) = $\boxed{۰٫۲۲}$

با توجه به مقادیر کاسبه شده توسط قاعده بیز و محاسبه likelihood تیم B احتمال بیشتری برای بردن در این بازی دارد.

* مخرج کسر بیز را برای سادگی حذف کردیم چون در هر دو کاسبه یک چیز است و مقدار رو تغییر نمی کند.

با توجه به محاسبات انجام شده مشخص است که تیم B برنده بازی خواهد بود بخاطر اینکه مقدار احتمال آن نسبت به تیم A بسیار بیشتر است. از قاعده بیز استفاده کردیم و محاسبات در عکس ها مشخص هستند.

سوال ٤

$$P(+) = \frac{\text{تعداد مثبت}}{\text{تعداد کل}} = 0/5 = \frac{5}{10}$$

$$P(-) = \frac{5}{10} = 0/5$$

سوال ۴

$$P(B=1|-) = \frac{P(B,-)}{P(-)} = \frac{\frac{2}{10}}{\frac{5}{10}} = \frac{2}{5} = 0/4 \quad P(B=0|-) = 0/6$$

$$P(A=1|-) = \frac{P(A,-)}{P(-)} = \frac{\frac{2}{10}}{\frac{5}{10}} = \frac{2}{5} = 0/4 \quad P(A=0|-) = 0/6$$

$$P(C=1|+) = \frac{P(C,+)}{P(+)} = \frac{\frac{4}{10}}{\frac{5}{10}} = \frac{4}{5} = 0/8 \quad P(C=0|+) = 0/2$$

$$P(A=1|+) = \frac{P(A,+)}{P(+)} = \frac{\frac{3}{10}}{\frac{5}{10}} = \frac{3}{5} = 0/6 \quad P(A=0|+) = 0/4$$

$$P(B=1|+) = \frac{P(B,+)}{P(+)} = \frac{\frac{1}{10}}{\frac{5}{10}} = \frac{1}{5} = 0/2 \quad P(B=0|+) = 0/8$$

$$P(C=1|-) = \frac{P(C,-)}{P(-)} = \frac{\frac{5}{10}}{\frac{5}{10}} = 1 \quad P(C=0|-) = 0$$

$$P(+|\underbrace{A=0,B=1,C=0}_{E}) = P(E|+) \times P(+) = \frac{4}{10} \times \frac{2}{10} \times \frac{2}{10} \times \frac{5}{10} = \frac{1}{125} = 0/008$$

$$P(-|\underbrace{A=0,B=1,C=0}_{E}) = P(E|-) \times P(-) = \frac{5}{10} \times 0 \times \frac{4}{10} \times \frac{6}{10} = 0$$

نمونه به کلاس "+" تعلق دارد.

با توجه به محاسبات انجام شده نمونه جدید به کلاس + تعلق دارد.

سوال ۵

الف)

k-mean به چند دلیل بسیار مناسب برای dataset بزرگ مناسب است:

کارایی محاسباتی: پیچیدگی زمانی k-mean از مرتبه O(N*K*D*I) است که n مجموعه داده ما و k تعداد خوشه ها و d بعد داده ها و i تعداد iteration ها است که بیانگر این است که الگوریتم رشد خطی دارد نسبت به مجموعه دادهها (n).

علاوه بر این به سبب سادگی عملیات ها در این الگوریتم که صرفا شامل محاسبه فاصله، تخصیص نقاط به خوشه ها و محاسبه میانگین برای مراکز خواهد بود و با وجود اینکه این یک الگوریتم iterative است چون محاسبات ساده است هزینه آن بسیار کم خواهد بود و سرعت پردازش بالا خواهد رفت.

ب)

روش های مختلفی برای انتخاب تعداد k پیشنهاد شده است شامل domain knowledge، Bayesian information criterion، Elbow method میشود.

Domain knowledge: در این روش ما اتکا میکنیم به دانش پیشین (prior) در مورد ساختار داده و مسئله ما.

Elbow Method: در این روش که برای مدل های گرافیکی است برای یک بازه ای از k ها ما الگوریتم را اجرا میکنیم و خطا را محاسبه میکنیم و به دنبال elbow point میگردیم، این نقطه جایی است که خطا کمترین مقدار ممکن است.

Shlhouette method:

در این روش شباهت بین نقاط داده با خوشه مربوط به خودشان را اندازه گیری میکنیم تا انسجام را بالا ببریم.

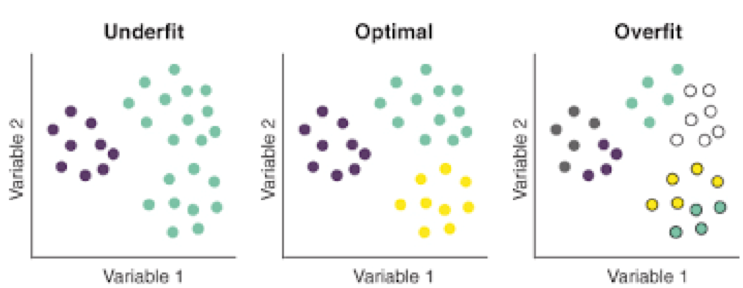
BIC: در این روش بیزینی یک نگاه آماری داریم که با استفاده از مدل های آماری Kیی را انتخاب میکند که از overfitting جلوگیری کند.

- ✓ تاثیر انتخاب k بر نتایج خوشه‌ها:

اگر K را خیلی کم بگیریم باعث میشود که داده های متفاوت با همدیگر مجبور شوند در یک خوشه قرار بگیرند و خوشه های معنی داری تولید نشوند. این امر باعث میشود خطای sum of squared بسیار زیاد شود و نتوانیم الگوی خوبی از داده ها استخراج کنیم.

اگر K را خیلی زیاد بگیریم باعث میشود خوشه ها بسیار خاص بشوند و نتوانیم ساختار های واقعی داده را بدست بیاوریم، یعنی که ما یک خوشه واقعی را به تعدادی زیر خوشه شکسته ایم، و باعث overfitting میشود. خطای sum of squared (WCSS) کم خواهد شد اما خوشه‌ها تعمیم پذیر نخواهند بود.

اگر مقدار K مناسب باشد باعث میشود که هم انسجام درونی خوشه ها بالاتر برود و نقاط شبیه به هم در یک خوشه قرار بگیرند. هم باعث میشود پراکندگی خوبی داشته باشیم و داده های متفاوت با همدیگر در خوشه های متفاوت قرار بگیرند. و در نهایت این امر باعث میشود خوشه های ما معنی دار باشند و الگوهای خوبی از آنها استخراج شود.



ث)

در الگوریتم k-mean و کلا مسائلی که داریم نسبت به convex بودن مسائل و ابر کروی مانند خوشه ها تاکید شده است به دلیل مسائل بهینه سازی که ممکن است از بهینه سازی خطی خارج شوند به این دلیل که مبنا را فاصله اقلیدوسی برای انتساب یک داده به یک خوشه در نظر گرفتیم و مرکز خوشه های در مرکز ابر کره قرار میگیرند.

در نتیجه برای اشکال پیچیده تر کارایی پایین میاد زیرا k-mean تلاش میکند آن ها را به صورت پیش فرض درستی که داشته است در بیاورد که باعث خطا میشود. یکی از روش های پیشنهاد شده در این راه ++k-mean خواهد بود که جلوتر به حل آن اشاره شده است.

سوال ٦

سوال ۶

$$P(F|C) = \frac{P(F,C)}{P(C)}$$

$$P(C) = \sum P(F|R)\,P(R|C)$$

$$P(F|C) = \sum P(F|R)\,P(R|C) = 0{,}1 \times 0{,}8 + 0{,}7 \times 0{,}2 = 0{,}22$$

مقدار نهایی 0.22 است همانطور که قابل مشاهده است.

سوال ۷

به طور کلی خوشه بندی سخت گیرانه و خوشه بندی فازی از چندین نظر بررسی کرد:

۱. عضویت:

خوشه بندی سخت گیرانه: در این رویکرد، هر داده دقیقا به یک خوشه تعلق دارد. این مقدار مطلق و انحصاری است. داده یا جز خوشه هست یا نیست. هیچ درجه عضویتی وجود ندارد. یکی از مثال های این رویکرد k-means است.

خوشه بندی فازی: یک داده میتواند به چندین خوشه به صورت همزمان تعلق داشته باشد، با مقادیر مختلف درجه عضویت نسبت به هر خوشه. این درجه عضویت مقداری بین ۰و ۱ خواهد بود. ۰ به معنای عدم عضویت در آن خوشه، ۱ به معنای عضویت کامل در آن خوشه و مقادیر بین ۰ و ۱ نشانگر عضویت نسبی به آن خوشه است.

۲. مرز خوشه‌ها:

HCM: در این رویکرد مرزهای خوشه ها با هم هیچ همپوشانی ندارند و مجزا هستند.

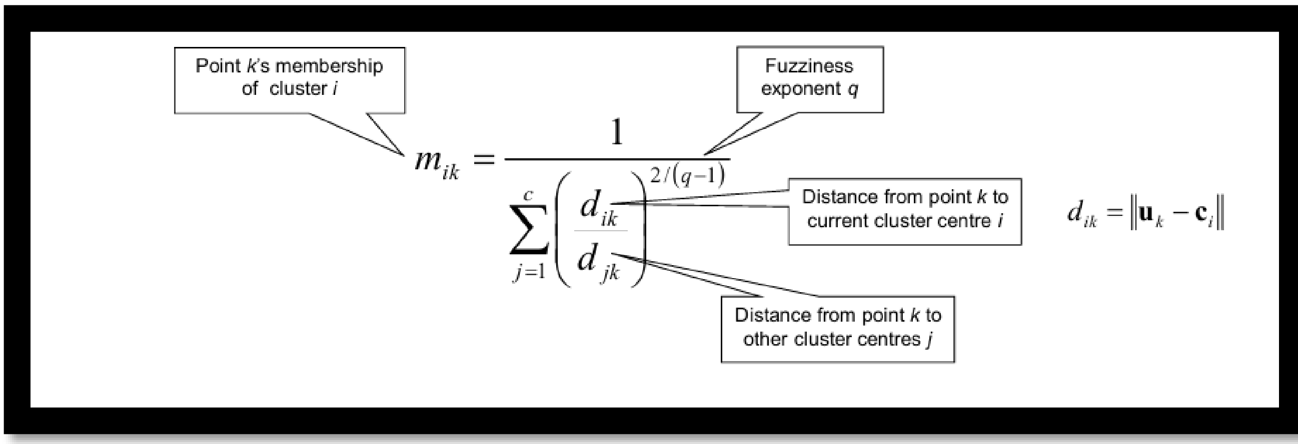
FCM: در این رویکرد مرز خوشه ها نرم و دارای همپوشانی بین خوشه های مختلف است. یک داده میتواند در یک ناحیه ای قرار بگیرد که خوشه های مختلف در آنجا همپوشانی داشته باشد و درجه عضویت بیانگر حضور آن داده در خوشه های متعلق به آن است.

به سبب همپوشانی بین خوشه های مختلف، FCM مستعد عدم قطعیت خواهد بود زیرا عضویت های نسبی به هر داده نسبت به خوشه های مختلف میدهد. یک داده در نزدیک مرز بین

۲ خوشه، درجه عضویت بالایی در هر ۲ میتواند داشته باشد و باعث عدم قطعیت و ابهام در تخصیص یک خوشه به یک داده میشود. اما در HCM به سبب اینکه اجازه عضویت نسبی نمیدهد حتی اگر یک داده در نزدیکی خوشه دیگر باشد، جز آن خوشه حساب نمیشود.

همچنین در رویکرد فازی ما نمایه فازی هم داریم که در سوال بعد به صورت کامل توضیح داده شده است.

فرمول درجه عضویت خوشه بندی فازی:



نسبت فاصله: اگر نسبت کسر کوچک باشد به این معنی است که داده به خوشه i نزدیک تر از خوشه j خواهد بود. اگر نسبت کسر بزرگ باشد به این معنی است که داده به خوشه j نزدیک تر است نسبت به خوشه i.

سوال ۸

نمایه فازی یکی از پارامتر ها در خوشه بندی فازی هست که درجه همپوشانی بین خوشه های مختلف را مشخص میکند یا به طور بهتر درجه عضویت بین خوشه های مختلف را مشخص میکند که در رویکرد FCM مورد استفاده قرار میگیرد که داده ها نه لزوما فقط به یک خوشه بلکه

میتوانند با یک درجه ای متعلق به تعدادی از خوشه ها باشند. این پارامتر تاثیر نسبت فاصله را کم یا زیاد میکند.

اگر q نزدیک به ۱ باشد تخصیص عضویت بسیار سخت گیرانه خواهد بود مانند k-means معمولی و مرز بین خوشه ها بسیار تند خواهد بود.

اگر q کمی بزرگتر از ۱ باشد، عضویت بیشتر فازی خواهد بود به این معنا که عضویت یک داده به چندین خوشه بسیار پر رنگ تر خواهد بود. مرز بین خوشه ها نرم تر و همپوشانی بیشتر خواهد شد.

اگر q به سمت بی نهایت میل کند، مقادیر عضویت برای همه خوشه ها تقریبا یکسان خواهد شد، به این معنی که فاصله عملا تاثیری ندارد و هر داده به همه خوشه ها به یک میزان مشابه عضویت دارد.

پس اگر پایین باشد، همپوشانی کمتر و hard خواهد بود. اگر زیاد باشد، همپوشانی بیشتر و مرز نرم تر خواهد بود و اگر خیلی بزرگ باشد عملا عضویت به یک میزان است.

$$1 < q < +\infty$$

$$1 \xleftarrow{\text{more crisp}} q \xrightarrow{\text{more fuzzy}} +\infty$$

سوال ۹

الف)

سوال ۹

الف)

$$WCSS = \sum_{i=1}^{k}\sum_{j=1}^{N} r_{ij} * \|x_j - \mu_i\|^{2}$$

$$\left.\begin{cases} r_{ij} = 1 & \text{if } i = \arg\min_c \|x_j - \mu_c\|^{2} \\ 0 & \text{otherwise} \end{cases}\right\} \text{Assignment}$$

update step

$$\mu_i = \frac{\left(\sum r_{ij} \times x_j\right)}{\sum r_{ij}}$$

از آنجایی که k-mean یک رویکرد iterative دارد و به صورت مکرر بروز رسانی میکند و سعی در کمینه سازی واریانس و افزایش bias درون هر خوشه را دارد پس از انجام تعداد محدودی سر انجام به یک بهینه محلی یا بهینه سراسری همگرا میشود.

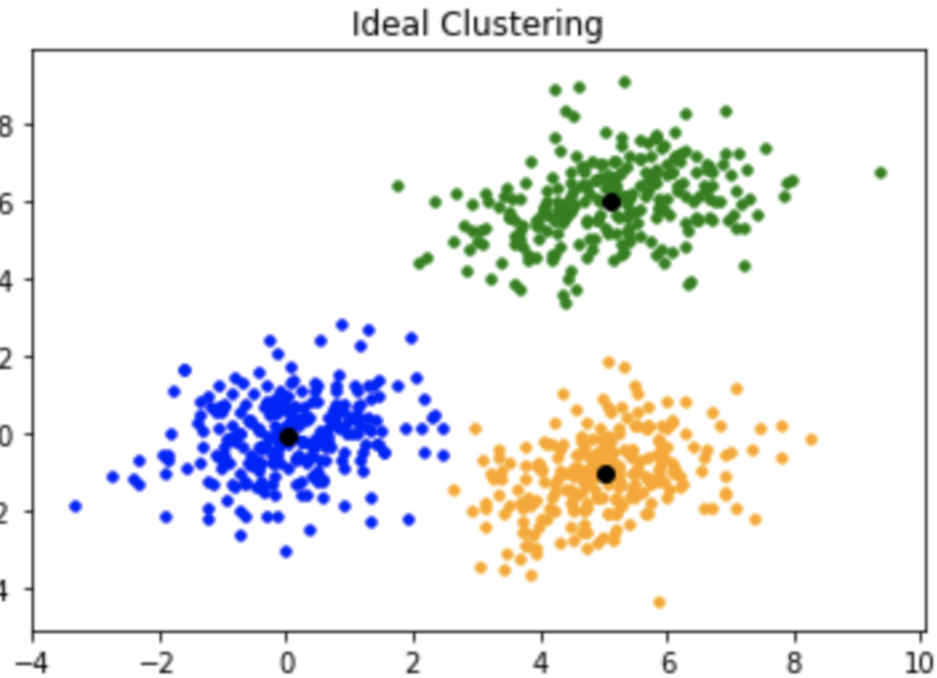
اگر واریانس در هر مرحله کم شود به این معنی است که bias در حال افزایش و واریانس درون خوشه ای در حال کاهش است یا متوقف شده است. البته این رویکرد چندین بار اجرا خواهد شد چون k-mean یک ماهیت تصادفی دارد. واریانس را همان مقدار WCSS میتواند تلقی کرد. همچنین این نکته مهم است که نقاط داده ما و خوشه های محدود هستند و به سمت بی نهایت میل نمیشوند.

ب)

مراکز خوشه ها در نهایت بیشتر احتمال دارد که در ناحیه ای که چگالی بیشتری دارد متمرکز میشوند، به چند دلیل:

الگوریتم k-mean و هر الگوریتمی که هدفشان کم کردن MSE است، چگالی تاثیر مستقیم بر آنها دارد. چونکه در نواحی که چگالی بالاتر است، داده ها نزدیک به یک دیگر هستند. قرار دادن مرکز یک خوشه در آن ناحیه باعث میشود که فواصل بین بسیاری از نقاط و مرکز کوچکتر شود در آن ناحیه، که این امر سبب میشود MSE به طور زیادی کاهش پیدا کند. اگر ناحیه ما چگالی پایین تری داشته باشد داده ها پراکنده هستند و مرکز یک خوشه فاصله بسیار زیادی تا نقاط دارد که باعث افزایش MSE میشود.

به همین دلیل k-mean به دنبال این است که MSE را کم کند به صورت طبیعی مراکز خوشه ها را در نواحی قرار میدهد که فاصله از بسیاری از نقاط به صورت همزمان کاهش پیدا کند، که به این معنی است که k-mean به سمت نواحی با چگالی بالاتر کشیده میشود.



پ)

مشکلی که در رابطه با k-mean عادی وجود دارد این است که خیلی به نقاط شروع حساس است و اگر نقاط شروع خوب توزیع نشده باشند ممکن است خوشه هایی که انتخاب میکند خیلی ضعیف و نامناسب باشند یا همگرایی کند صورت بگیرد. همچنین همانطور که اشاره شد k-mean نقاط ابتدایی مراکز را معمولا به صورت تصادفی از داده ها انتخاب میکند.

در نسخه تغییر داده شده k-mean که به k-mean++ معروف است، الگوریتم رویکرد بهتری را برای انتخاب اولیه مراکز در نظر میگیرد.

در این نسخه به سبب d^2، نقاطی که بسیار دورتر از مراکز فعلی ما هستند احتمال بیشتری برای انتخاب شدن به عنوان مراکز جدید را دارند. و توان ۲ این تاثیر را بسیار تقویت میکند.

این نسخه به صورت کلی همگرایی سریعتری نسبت به نسخه معمولی k-mean دارد به سبب اینکه با انتخاب مراکز بهتری که خوب توزیع شده اند، خوشه های بهتری میسازد و iteration های کمتری برای رسیدن به همگرایی نیاز هست. همچنین در نتایج خوشه بندی هم بسیار خوب عمل میکنند و احتمال گیر کردن در local optima را کاهش میدهند.

اما سرعت این نسخه نسبت به نسخه اصلی k-mean کند تر است، به سبب محاسبه توزیع احتمالی در هر iteration.