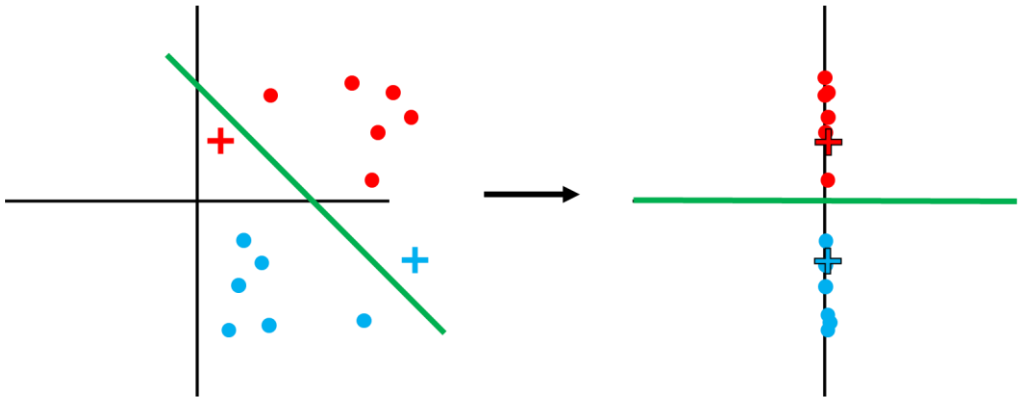
Sharif University of Technology
Department of Computer Engineering
June 2022

Hadis Ahmadian
400211524
ML for Bioinformatics Assignment #3

Question ONE .....

.1

(آ)



دیتاست سمت چپ را در نظر بگیرید نقاط قرمز متعلق به کلاس 1 و نقاط آبی متعلق به کلاس 2 هستند. داده های به شکل دایره داده های آموزش و داده های + شکل داده های تست هستند.

اگر برای داده ها یک مرز تصمیم در 2 بعد پیدا کنیم انتظار داریم دسته بند ما مانند خط سبز در شکل سمت چپ شود چون به نظر دسته بند خوبی است(از این جهت که فاصله اش با هردو کلاس تقریبا یکسان است و این فاصله را بیشینه کرده است.)

با این حال این دسته بند نمیتواند داده های تست را درست دسته بندی کند.

اگر داده ها را روی محور عمودی مپ کنیم یعنی درواقع فقط همین بعد آن را درنظر بگیریم میتوان دید یک مرز تصمیم دیگر(خط افقی سبز) برای داده پیدا خواهد شد که دیگر مشکل دسته بندی غلط داده های تست را نخواهد داشت.

میتوان دید که دقت دسته بند روی داده های تست بعد از کاهش بعد بیشتر است و داده های تست درست دسته بندی شده اند.

(ب) دو مرحله ی Evaluation و subset generation.

Evaluation: هدف آن معیاری برای سنجش یک زیر مجموعه ی کاندید از ویژگی هاست.

متد های Evaluation :

Feature ranking: به هر ویژگی طبق یک معیار و استانداردی یک رنک داده میشود و اگر رنک آن به آستانه ی مشخصی نرسد، از مجموعه ویژگی ها حذف میشود. مثال هایی برای این متد : **Pearson correlation, mutual information** و **information gain**.

Probabilistic distance: محاسبه ی فاصله ی احتمالاتی بین احتمال شرطی دوکلاس یعنی $p(x|C1)$ و $p(x|C2)$. مثال هایی برای این متد : **Chernoff dissimilarity measure**

Subset generation: هدف آن تولید یک زیر مجموعه برای سنجیده شدن توسط **Evaluation** است.

متد های Subset generation:

Complete search: این متد ها تضمین میکنند که به جواب بهینه (طبق معیار انتخاب شده برای **Evaluation**) خواهند رسید. برای مثال **exhaustive search** با جستجو و سنجیدن کل فضای حالت یعنی بین تمام زیر مجموعه های ممکن از ویژگی ها، پاسخ بهینه را میابد و بنابراین یک جستجوی کامل است.

Best individual N: یک رویکرد ساده این است که به هر ویژگی یک رنک (طبق معیارانتخاب شده برای **Evaluation**) داده شود و **N** بهترین آن ها به عنوان زیرمجموعه ی جواب برگردد.

.2

$z_1 = \omega_1^T x$

$Var(z_1) = E[(\omega_1^T x - \omega_1^T \mu)^2] = E[(\omega_1^T x - \omega_1^T \mu)(\omega_1^T x - \omega_1^T \mu)^T]$

$= E[\omega_1^T (x-\mu)(x-\mu)^T \omega_1] =$

$\omega_1^T E[(x-\mu)(x-\mu)^T]\omega_1 = \omega_1^T \Sigma \omega_1$

Cov matrix

$\Rightarrow Var(z_1) = \omega_1^T \Sigma \omega_1$ (1)

داده‌ی map شده

چون در PCA خطی ما می‌خواهیم این Var بیشینه شود:

$\text{argmax } \omega_1^T \Sigma \omega_1 \quad \text{Subject to } \omega^T \omega = 1$

نوشتن به فرم لاگرانژ ها:

$\underset{\omega_1}{\text{maximize}} \ \omega_1^T \Sigma \omega_1 - \alpha(\omega_1^T \omega_1 - 1)$

$\partial / \partial \omega_1 = 0 \Rightarrow 2\Sigma \omega_1 = 2\alpha \omega_1 \Rightarrow \Sigma \omega_1 = \alpha \omega_1$ (2)

مشتق گیری بر حسب ω_1 و مساوی صفر قرار دادن برای max شدن

بنابراین ω_1 بردار ویژه‌ی ماتریس کوواریانس (Σ) است و α نیز مقدار ویژه‌ی آن است.

(1) و (2) طبق $\Rightarrow Var(z_1) = \omega_1^T \Sigma \omega_1 = \alpha \omega_1^T \omega_1 = \alpha$

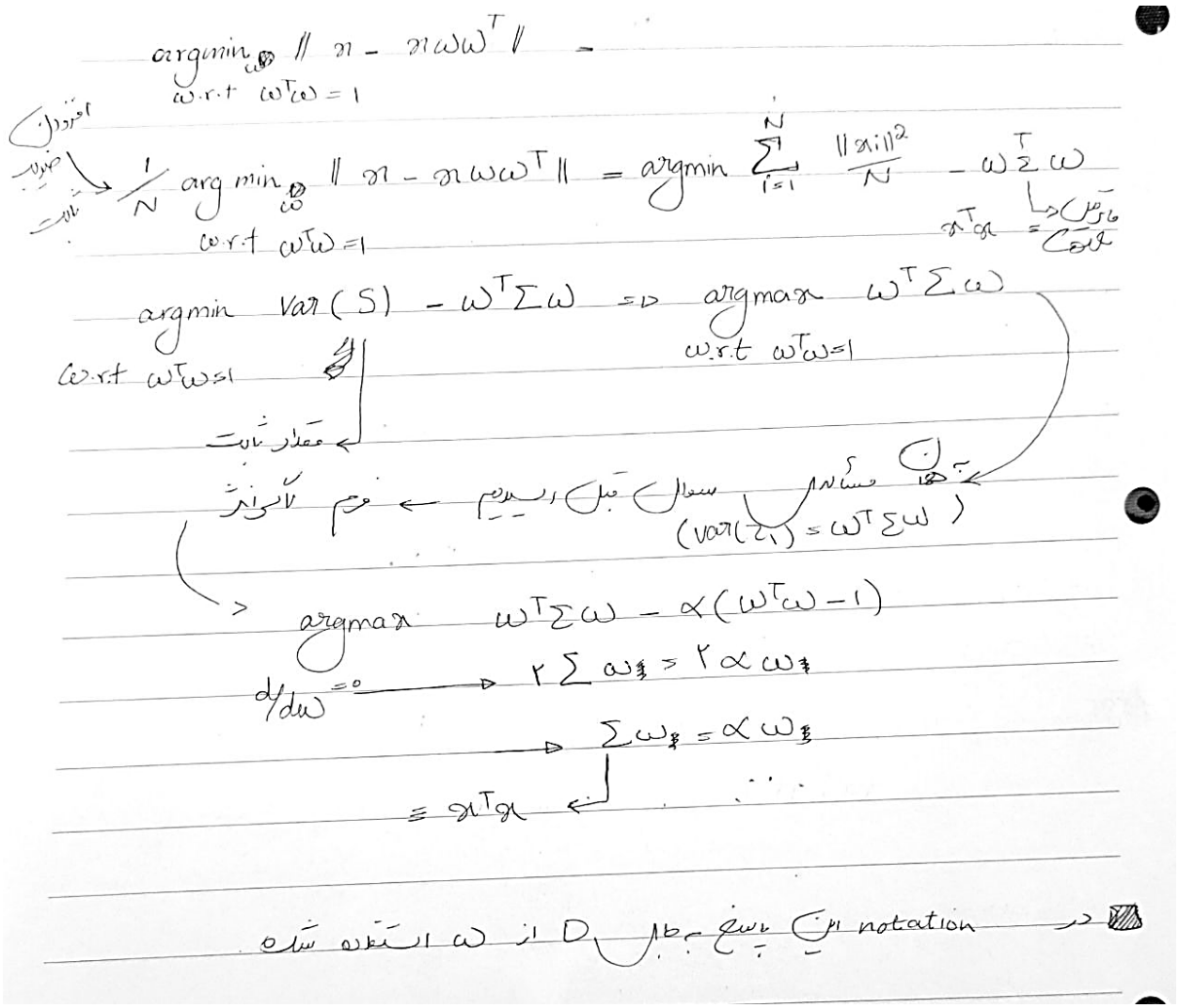
$Var(z_1) = \alpha$

پس واریانس z_1 برابر است با مقدار ویژه‌ی ماتریس Cov (Σ) و

z_1 هم حاصل نگاشت داده روی بردار ویژه‌ی Σ است ($z_1 = \omega_1^T x$)

ثابت شد بردار ویژه‌ی Σ است

.3



.4

(آ) منظور این است که ما فرض میکنیم یک سری فاکتورهای غیرقابل مشاهده و نهفته وجود دارند که ترکیب این فاکتور ها **x** را حاصل میشود. پس در واقع هدف ما این هست که **dependency** ویژگی هارا با ترکیبی از فاکتور ها که تعداد آن ها کمتر از ویژگی های اصلی است بیان کنیم. برای مثال ممکن است چندین ویژگی با همبستگی بسیار بالا داشته باشیم که این ها همه میتوانند حالت های مختلف یک فاکتور باشند. کاربرد آن علاوه بر کاهش بعد میتواند **knowledge extraction** و بیان متغیرها با تعداد فاکتور های کمتر باشد.

(ب)

Multidimensional Scaling: MDS سعی میکند تمام نقاط را به یک بعد کمتر مپ کند به گونه ای که فاصله ی دو به دوی آن ها تا حد ممکن تغییر نکند. این فاصله میتواند فاصله ی اقلیدسی یا هر معیار فاصله ی دیگری باشد. در واقع MDS سعی میکند تابع زیر را کمینه کند که در آن نقاط X1….Xn به نقاط Z1….Zn در ابعاد کمتری مپ شده اند و dij فاصله ی دو به دوی Xi هاست و d^ij فاصله ی دو به دوی zi هاست و درواقع تابع دارد سعی میکند فواصل دو به دوی نقاط در فضای جدید تا حد امکان شبیه نقاط اولیه باشد و مجموع مجذورات تفاوت این فاصله ها کمینه باشد(فاصله در اینجا اقلیدسی است).

$$\min_z \sum_{i=1}^{N} \sum_{j=1}^{N} \left(d_{ij} - \hat{d}_{ij} \right)^2$$

فاصله ها فواصل اقلیدسی هستند و تابع هدف MDS به فرم زیر کاهش میابد (reduce میشود) که در آن Xi یک نقطه در فضای اصلی و Zi معادل آن نقطه در فضای جدید است. X^TX نیز درواقع ماتریس Gram است.

$$\min_z \sum_{i=1}^{N} \sum_{j=1}^{N} \left(x_i^T x_j - z_i^T z_j \right)^2$$

Locally Linear Embedding: در این روش ساختار های غیرخطی در فضای سراسری بر اساس فیت های خطی در نواحی محلی به دست می آید. ایده ی اصلی LLE این است که هر نقطه جمع وزن داری از همسایه هایش است (تعریف همسایه میتواند بر اساس یک فاصله ی مشخص یا تعداد مشخصی نقاط نزدیک باشد). در مقوله ی کاهش بعد، این روش در این مورد قابل استفاده است که وزن های به دست آمده، ویژگی های هندسی ذاتی نقاط را منعکس میکنند بنابراین در ابعاد دیگر نیز معتبر خواهند بود.

در مرحله ی اول LLE ما میخواهیم وزن ها را به گونه ای پیدا کنیم که برای هر نقطه، جمع وزن دار همسایه هایش تا حد امکان به خود نقطه نزدیک باشد. برای هر نقطه x^r همسایه های آن را با نماد X_r^s نشان میدهیم. Wrs هم وزنی است که در همسایه ی s ام Xr ضرب میشود و باید مجذور فاصله ی این مجموع وزن دار از نقطه ی اصلی حساب شود. هدف LLE این است که مجموع این فاصله ها را کمینه کند یعنی وزن ها به گونه ای باشند که مجموع وزن دار همسایه های هر نقطه تا حد امکان به آن نقطه نزدیک باشد. و نکته ی دیگر هم اینکه برای هر نقطه باید مجموع وزن همسایگانش 1 باشد.

$$E[W|x] = \sum_{r=1}^{N} \left\| x^r - \sum_s w_{rs} x_{(r)}^s \right\|^2$$

$$\text{subject to } \sum_s w_{rs} = 1$$

سپس در مرحله ی دوم این وزن ها فیکس میشوند و LLE تلاش میکند با کمینه کردن تابع زیر مختصات Y(که در ابعاد جدید است) را به گونه ای بیابد که طبق همان وزن ها، این الگوهای خطی محلی (یعنی درواقع آن هندسه ی ذاتی نقاط) حفظ شوند. به این منظور تابع زیر را کمینه میکند که در آن Y نقاط در ابعاد جدید است، Wrs وزن های محاسبه شده در مرحله ی قبل است و Y_r^s همسایه های نقطه ی r هستند. با کمینه کردن تابع زیر میخواهیم شرط (مجموع به ازای هر نقطه ی مجذور فاصله ی مجموع وزن دار همسایه ها از نقطه ی اصلی کمینه شود) در فضای جدید نیز برقرار باشند. در ضمن شرط Cov(Y) = I and E [Y] = 0 نیز باید برقرار باشد.

$$\mathbb{E}[Y|W] = \sum_{r=1}^{N} \left\| y^r - \sum_s w_{rs} y^s_{(r)} \right\|^2$$

$$\text{in such a way that } Cov(Y) = I \text{ and } \mathbb{E}[Y] = 0$$

Isomap: رویکرد **isomap** مانند MDS است با این تفاوت که به جای فاصله ی اقلیدسی از نوع دیگری از فاصله استفاده میکند و در واقع یک **generalization** غیرخطی از MDS است. ایزومپ ابتدا فواصل **geodesic** را محاسبه میکند و سپس MDS را روی این فواصل اجرا میکند. فاصله ی **geodesic** در واقع کوتاه ترین مسیر روی سطح منحنی است گویی که سطح **flat** باشد.

برای محاسبه ی این نوع فاصله میتوان از الگوریتم هایی مانند فلویدوارشال استفاده کرد.

این الگوریتم ابتدا روی فضای منحنی برای هرنقطه نقاطی را میابد که از لحاظ فاصله ی اقلیدسی فاصله ی کمتری نسبت به یک آستانه داشته باشند (d(i,j)<e) سپس این نقاط را با یال به یکدیگر متصل میکند. در نتیجه یک گراف روی سطح منحنی ایجاد میشود. برای هر دو نقطه یعنی هر دو راس در این گراف کوتاه ترین مسیر محاسبه میشود و به شکل dG(i,j) ذخیره میشود. حال ایزومپ عینا مانند MDS عمل میکند با این تفاوت که معیار فاصله ها را dG(i,j) در نظر میگیرد. همانطور که در MDS توضیح داده شد، نقاط X1….Xn به نقاط Z1….Zn در ابعاد کمتری مپ شده اند ودر ایزومپ dG(ij) فاصله ی دو به دوی Xi ها در گراف است و d^G (ij) فاصله ی دو به دوی zi هاست و درواقع تابع دارد سعی میکند فواصل **geodesic** دو به دوی نقاط در فضای جدید تا حد امکان شبیه نقاط اولیه باشد و مجموع مجذورات تفاوت این فاصله ها کمینه باشد(فاصله در اینجا **geodesic** است که با مدل سازی به گراف و محاسبه ی کوتاه ترین مسیر بین هر دو راس به دست آمد).

$$\min_z \sum_{i=1}^{N} \sum_{j=1}^{N} \left(d_G(i,j) - \hat{d}_G(i,j) \right)^2$$

■

Question TWO .....

.1

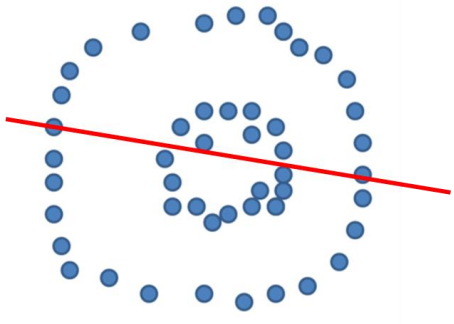
در kmeans ما میخواهیم برای k کلاستر، $(x-\mu_k)^2$ را کمینه کنیم که x داده های ما هستند و m مرکز هر کلاستر است.

در GM با k کلاستر و ماتریس کواریانس به شکل گفته شده، هدف کمینه شدن مقدار $\frac{(x-\mu_k)^2}{\sigma^2}$ است. زیرا نوع ماتریس کواریانس spherical است و این یعنی تمام خوشه ها دقیقا دایره هستند (صورت کسر) که این اتفاق دقیقا در فاصله ی اقلیدسی هم میافتد (یک دایره نقاطی هستند که فاصله ی اقلیدسیشان نسبت به مرکز – در اینجا مرکز خوشه- یکسان است)

.2

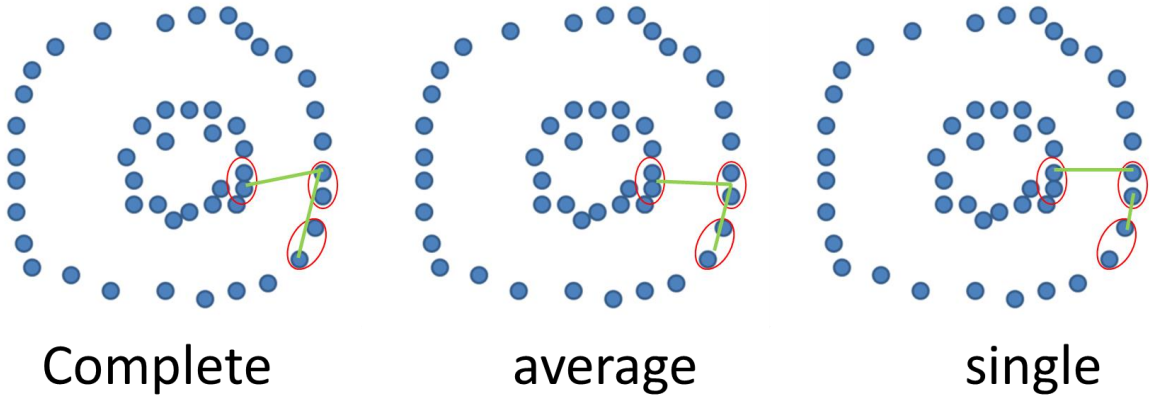
(آ) هر دو خوشه ها را با مرزی به این شکل (و نه با یک مرز دایره ای) از هم جدا میکنند دلیل آن این است که k-means هدفش پیدا کردن k=2 خوشه خواهد بود به گونه ای که دو مرکز خوشه پیدا کند به طوری که مجموع فاصله ی اجزای هر خوشه از مرکز آن خوشه کمینه شود به طور کلی یک سری مرکز خوشه و دایره هایی در اطراف آن در فضا رسم میکند (بر اساس فاصله ی اقلیدسی) به عنوان خوشه ها و بنابراین نمیتواند دو خوشه که مانند شکل زیر درون هم قرار دارند را دسته بندی کند زیرا مرکز این دو خیلی به هم نزدیک هستند ولی اینگونه نیست که هرکدام در دو دایره ی جدا نسبت به مرکز خود قرار بگیرند بلکه تفاوت آنها در فاصله شان از مرکز حدودا مشترکشان است و این نوع خوشه ها را kmeans به شکل زیر دسته بندی میکند.

مدل EM هم با 2 توزیع گوسی، دقیقا همین رفتار را خواهد داشت و مانند توضیحات بالا میباشد و کلا Kmeans خود حالت خاصی از EM با توزیع گوسی است.

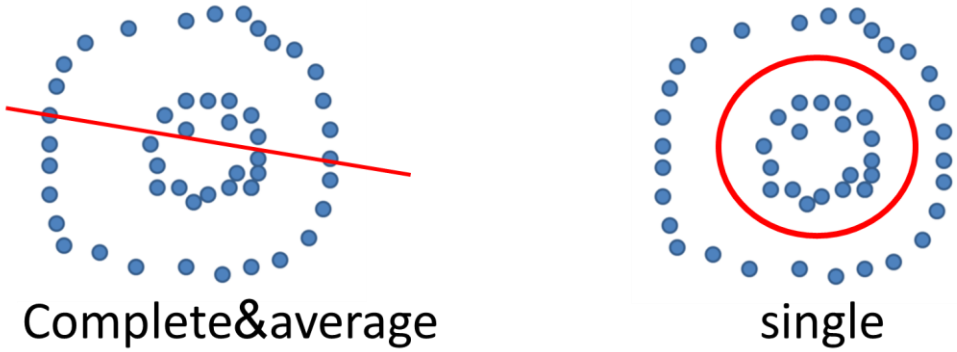


(ب) در ابتدا که هر نقطه یک کلاستر جدا گانه است، در هر حالت نقاط نزدیک تر به هم یک کلایتر میشوند حالا برای مراحل بالاتر، در مورد single-link میتوان مطمین بود که خوشه هایی که در ادامه با هم merge میشوند خوشه های تشکیل دهنده ی هر حلقه هستند پس دو خوشه ی به شکل دو حلقه از هم جدا خواهند شد. اما در دو حالت دیگر همانطور که در مثال زیر مشخص است مثلا در مورد average ممکن است مرکز دو خوشه در یک حلقه از هم دورتر باشد تا مرکز دو خوشه در دو حلقه ی متفاوت هم چنین در

حالت **complete** نیز ممکن است ماکسیمم فاصله ی دو خوشه در یک حلقه از هم بیشتر باشد تا بیشترین فاصله ی دو خوشه در دو حلقه ی متفاوت بنابراین دو خوشه به شکل دو حلقه از هم جدا نخواهند شد.



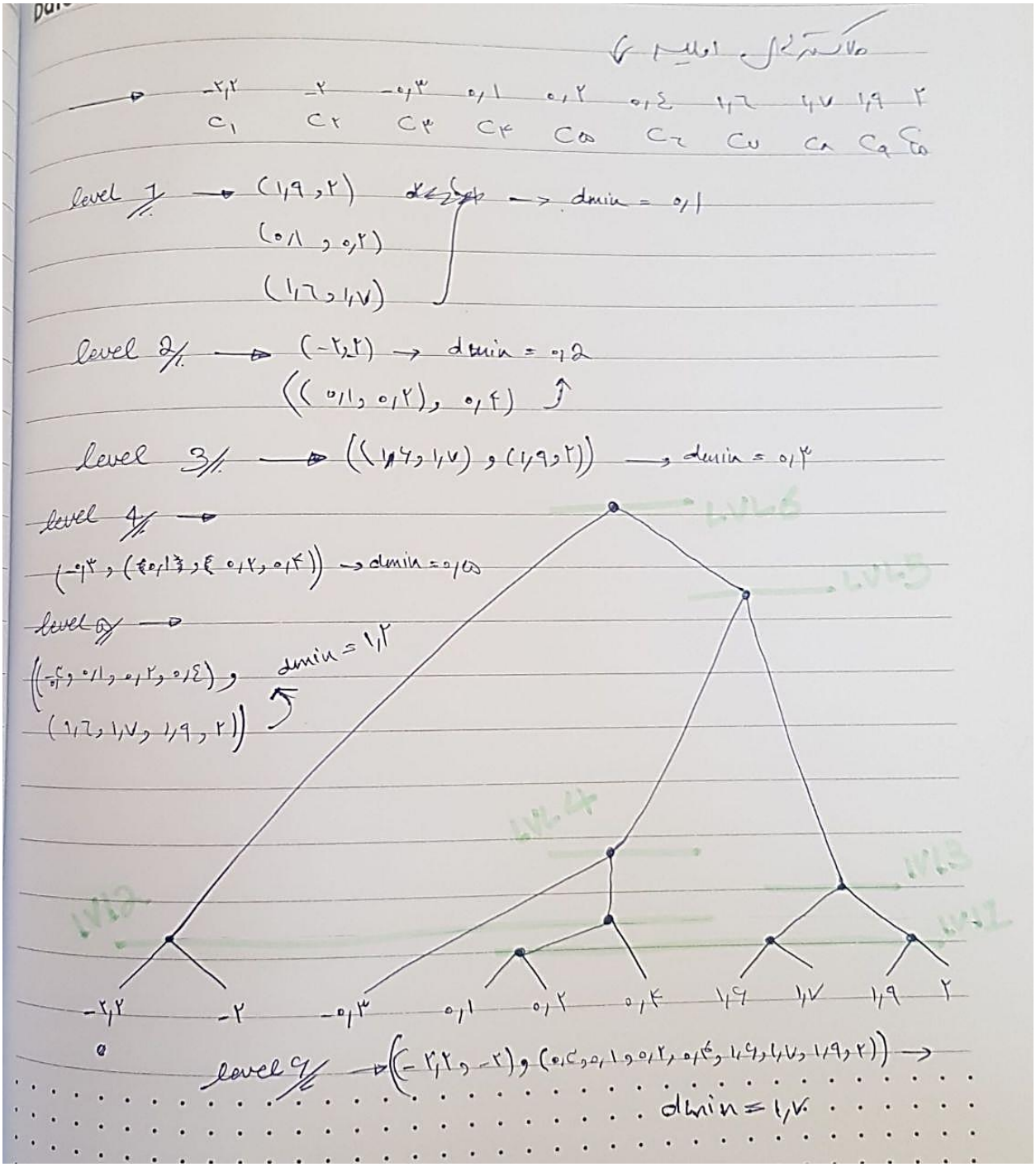
در نتیجه نتیجه ی نهایی کلاسترینگ حدودا به شکل زیر خواهد بود:



.3

وقتی میخواهیم کلاستر هایی که طبق معیار گفته شده بیشترین شباهت را به هم دارند انتخاب کنیم در واقع میخواهیم کلاسترهایی را انتخاب کنیم که کمترین **dmin** را باهم دارند بنابراین با این روش، ابتدا تمام نقاط را یک کلاستر در نظر میگیریم سپس آنهایی که

کمترین **dmin** را باهم دارند با هم **merge** میکنیم و در **level** بعدی نیز همین کار را تکرار میکنیم تا به درخت نهایی برسیم. کلاستر هایی که در هر **level** انتخاب میشوند و **dmin** آنها نسبت به هم و هم چنین نتیجه ی نهایی در ادامه آمده است.



■

Question THREE.....

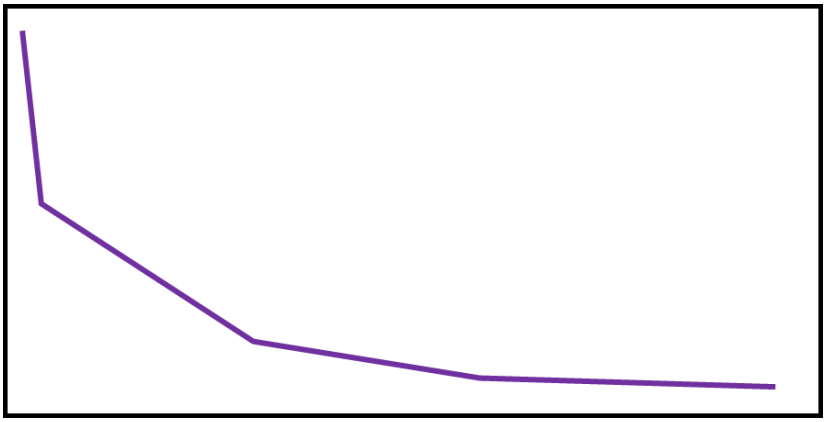
.1

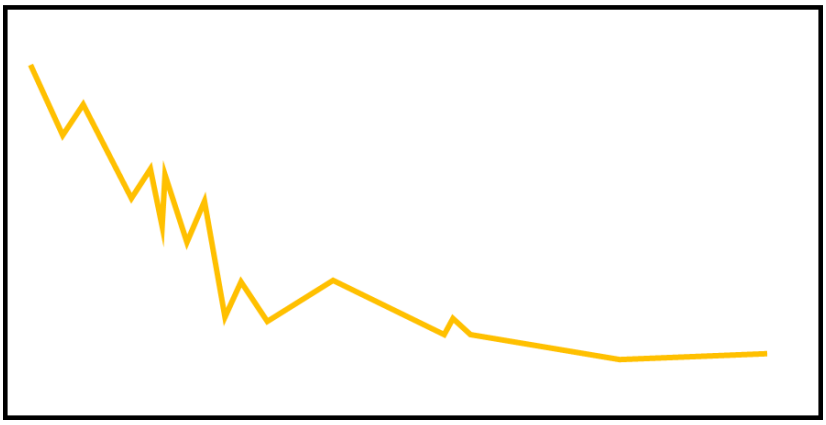
(آ) پرسپترون یک الگوریتم **iterative** است که از یک نقطه ی شروع به سمت یک اوپتیمال محلی برای تابع هزینه حرکت میکند بنابراین نقطه ی شروع در آن تاثیر زیادی دارد و داده هایی که در ابتدا دیده میشوند تاثیرگذار تر هستند از داده هایی که در ایپاک های آخر دیده میشوند اگر چند تک داده ی اول **outlier** باشند، چون پرسپترون در هر ایپاک فقط یک داده را میبیند به سمت داداه های اولیه ی **outlier** بایاس میشود. به همین علت بهتر است در کابرد هایی که به نقطه ی شروع حساس هستد چندین بار الگوریتم را از نقطه های شروع متفاوت آغاز کرد.

(ب) در حالت **Minibatch** چون پرسپترون در هر ایپاک چندین داده میبنید و برای تصمیم جهت حرکت در فضای حالت، به گرادیان تمام آن بچ توجه میکند این باعث میشود حرکت نرم تری داشته باشد(چون یک داده ی **outlier** در یک بچ تاثیر کمتری دارد و اثر سایر داده های سالم آن را تا حد خوبی فیلتر میکند) و با یک شیب نسبتا نرم خطا در هر ایپاک کاهش میابد.

در حالت تک داده پرسپترون در هر ایپاک با توجه به گرادیان تنها یک داده جهت حرکت در فضای حالت را مشخص میکند این باعث میشود اگر داده **outlier** باشد یا نویز زیادی داشته باشد و حرکت تنها با اتوجه به آن صورت بگیرد، خطا روی سایر داداه ها افزایش یابد پس در این حالت حرکت خیلی بیشتر نویز دارد و مانند حالت قبل یکنواخت نیست.

نمونه ای از مقدار تابع خطا در هر ایپاک برای دو حالت فوق:





2.

نکات مهم:

در این پاسخ $z_{2\_final} = \sigma(o_2)$ & $z_{1_{final}} = \sigma(o_1)$ & $z_2 = \sigma(h_2)$ & $z_1 = \sigma(h_1)$ که σ تابع فعال ساز سیگموید است.

هم چنین در مشتق گیری ها تمام محسابات تا 3 رقم اعشار انجام شده است.

علاوه بر لایه ی **hidden**، برای لایه ی خروجی نیز تابع سیگموید در نظر گرفته شده است.

$h_1 = w_1 x_1 + w_2 x_2 + w_3 x_3 + b_1$

$= 0.1 \times 1 + 0.4 \times 4 + 0.8 \times 5 + 0.6 = 4.4$

$Z_1 = \sigma(h_1) = \frac{1}{1+e^{-4.4}} = 0.986$

$h_2 = w_4 x_1 + w_5 x_2 + w_6 x_3 + b_1$

$= 0.2 + 1.2 + 1.2 + 0.5 = 3.1$

$Z_2 = \sigma(h_2) = \frac{1}{1+e^{-3.1}} = 0.957$

$O_1 = w_7 Z_1 + w_8 Z_2 + b_2 =$

$0.7 \times 0.986 + 0.9 \times 0.957 + 0.05 = 2.0857$

$\sigma(O_1) = \frac{1}{1+e^{-2.0857}} = 0.889 \rightarrow Z_{1\text{-Final}}$

$O_2 = w_9 Z_1 + w_{10} Z_2 + b_2 =$

$0.8 \times 0.986 + 0.1 \times 0.957 + 0.05 = 1.4488$

$\sigma(O_2) = \frac{1}{1+e^{-1.4488}} = 0.810 \rightarrow Z_{2\text{-Final}}$

$SSE = (t_1 - Z_{1\text{-Final}})^2 + (t_2 - Z_{2\text{-Final}})^2$

$= (0.1 - 0.889)^2 + (0.05 - 0.8)^2 =$

$0.622 + 0.562 = 1.184$

$\frac{\partial E}{\partial Z_{1\text{-Final}}} = 2 \times (-1)(t_1 - Z_{1\text{-Final}}) = -2(t_1 - Z_{1\text{-Final}})$

$= -2(-0.789) = 1.578$

$\frac{\partial E}{\partial Z_{2\text{-Final}}} = 2(-1)(t_2 - Z_{2\text{-Final}}) = -2(t_2 - Z_{2\text{-Final}})$

$= -2(-0.75) = 1.5$

$$\frac{\partial E}{\partial w_7} = \frac{\partial E}{\partial z_{1\text{-final}}} \times \frac{\partial z_{1\text{-final}}}{\partial o_1} \times \frac{\partial o_1}{\partial w_7}$$

$$\frac{\partial E}{\partial z_{1\text{-final}}} = 1.0758$$

$$\frac{\partial z_{1\text{-final}}}{\partial o_1} = \sigma'(o_1) = \frac{e^{-o_1}}{(1+e^{-o_1})^2} = 0.0948$$

$$\frac{\partial o_1}{\partial w_7} = z_1 = 0.9867$$

$dE/dw_7 =$
$1.0758 \times 0.0948 \times$
$\times 0.9867 =$
0.1007

$$\partial E/\partial w_9 = \frac{\partial E}{\partial z_{1\text{-final}}} \times \frac{\partial z_{1\text{-final}}}{\partial o_1} \times \frac{\partial o_1}{\partial w_9}$$

$$\partial z_{1\text{-final}}/\partial w_9 = z_2 = 0.9930$$

$dE/dw_9 =$
$1.0758 \times 0.0948 \times$
$0.9930 = 0.1012$

$$\partial E/\partial w_8 = \frac{\partial E}{\partial z_{2\text{-final}}} \times \frac{\partial z_{2\text{final}}}{\partial o_2} \times \frac{\partial o_2}{\partial w_8}$$

$$\frac{\partial z_{2\text{final}}}{\partial o_2} = \sigma'(o_2) = \sigma'(1.2888) = 0.1049$$

$$\partial o_2/\partial w_8 = z_1 = 0.9867$$

$dE/dw_8 =$
$1.0[illegible] \times$
$0.1049 \times$
$0.9867 =$
0.1035

$$\partial E/\partial w_{10} = \frac{\partial E}{\partial z_{2\text{final}}} \times \frac{\partial z_{2\text{final}}}{\partial o_2} \times \frac{\partial o_2}{\partial w_{10}}$$

$$\frac{\partial o_2}{\partial w_{10}} = z_2 = 0.9930$$

$1.0[illegible] \times$
$0.1049 \times$
$0.9930 = 0.1042$

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial z_1} \times \frac{\partial z_1}{\partial h_1} \times \frac{\partial h_1}{\partial w_1}$$ = ۰٫۲۹۸ × ۰٫۰۱۳ × ۱ = ۰٫۰۰۴

$$\frac{\partial E}{\partial z_1} = \frac{\partial E_1}{\partial z_1} + \frac{\partial E_2}{\partial z_1}$$ = (۰٫۱۰۸ + ۰٫۱۹۰) = ۰٫۲۹۸

$$\frac{\partial E_1}{\partial z_1} = \frac{\partial E_1}{\partial O_1} \times \frac{\partial O_1}{\partial z_1} = \frac{\partial E_1}{\partial z_{1final}} \times \frac{\partial z_{1final}}{\partial O_1} \times \frac{\partial O_1}{\partial z_1}$$

($\frac{\partial E_1}{\partial z_{1final}}$ ← از قبل محاسبه شده؛ $\frac{\partial z_{1final}}{\partial O_1}$ → $\sigma'(O_1)$؛ $\frac{\partial O_1}{\partial z_1}$ → w_1)

۱٫۵۷۸ × ۰٫۰۹۸ × ۰٫۷ = ۰٫۱۰۸

$$\frac{\partial E_2}{\partial z_1} = \frac{\partial E_2}{\partial O_2} \times \frac{\partial O_2}{\partial z_1} = \frac{\partial E_2}{\partial z_{2final}} \times \frac{\partial z_{2final}}{\partial O_2} \times \frac{\partial O_2}{\partial z_1}$$

($\frac{\partial z_{2final}}{\partial O_2}$ → $\sigma'(O_2)$؛ $\frac{\partial O_2}{\partial z_1}$ → w_8)

۱٫۵ × ۰٫۱۵۹ × ۰٫۸ = ۰٫۱۹۰

$$\frac{\partial z_1}{\partial h_1} = \sigma'(h_1) = \sigma'(4.3)$$ = ۰٫۰۱۳

$$\frac{\partial E}{\partial w_4} = \frac{\partial h_1}{\partial w_4} \times \frac{\partial E}{\partial z_1} \times \frac{\partial z_1}{\partial h_1}$$ = ۴ (x۲) × ۰٫۲۹۸ × ۰٫۰۱۳ = ۰٫۰۱۵

$$\frac{\partial E}{\partial w_5} = \frac{\partial h_1}{\partial w_5} \times \frac{\partial E}{\partial z_1} \times \frac{dz_1}{dh_1}$$ = ۵ (x۳) × ۰٫۲۹۸ × ۰٫۰۱۳ = ۰٫۰۱۹

$\partial E/\partial w_{r} = \frac{\partial E}{\partial z_r} \times \frac{\partial z_r}{\partial h_r} \times \partial h_r/\partial w_r$ (Local) $= 0.192 \times 0.004 \times 1 = 0.000$

$\partial z_r/\partial h_r = \sigma'(h_r) = \sigma'(0.1^{r}) = 0.004$

$\frac{\partial E}{\partial z_r} = \partial E_1/\partial z_r + \partial E_r/\partial z_r = 0.1^{r}9 + 0.00^{r} = 0.192$

$\partial E_1/\partial z_r = \partial E_1/\partial o_1 \times \partial o_1/\partial z_r = \partial E_1/\partial z_{1\,final} \times \frac{\partial z_{1\,final}}{\partial o_1} \times \partial o_1/\partial z_r$ ($\to w_9$)

$1.0VA \times 0.09A \times 0.9 = 0.1^{r}9$

$\partial E_r/\partial z_r = \partial E_r/\partial o_r \times \partial o_r/\partial z_r = \frac{\partial E_r}{\partial z_{r\,final}} \times \frac{\partial z_{r\,final}}{\partial o_r} \times \partial o_r/\partial z_r$ ($\to w_{10}$)

$1.0 \times 0.1^{r}9 \times 0.1 =$

($\to x_r$)

$\partial E/\partial w_4 = \partial h_r/\partial w_4 \times \frac{dE}{dz_r} \times \frac{dz_r}{dh_r} = 5 \times 0.192 \times 0.004 = 0.00^{r}$

$\partial E/\partial w_3 = \partial h_r/\partial w_3 \times \frac{dE}{dz_r} \times \frac{dz_r}{dh_r} = 5 \times 0.192 \times 0.004 =$

($\to x_r$)

0.00^{r}

$$\frac{dE}{db_2} = \frac{dE_1}{db_2} + \frac{dE_2}{db_2} =$$

$$\frac{dZ_{1final}}{dO_1} \times \frac{dO_1}{db_2} + \frac{dZ_{2final}}{dO_2} \times \frac{dO_2}{db_2}$$

$$= \sigma'(O_1) + \sigma'(O_2) = ۰/۰۹۸ + ۰/۱۵۹ = ۰/۲۵۷$$

$$dE/db_1 = \frac{dE}{dz_1} \times \frac{dz_1}{dh_1} \times \frac{dh_1}{db_1} + \frac{dE}{dz_2} \times \frac{dz_2}{dh_2} \times \frac{dh_2}{db_1}$$

$$۰/۴۹۸ \times ۰/۰۱۳ \times ۱ + ۰/۱۶۲ \times ۰/۰۰۴ \times ۱ = ۰/۰۰۴$$

→ New values for Bias.

$$b_1 = b_1 - ۰/۰۱ \times dE/db_1 = ۰/۵ - ۰/۰۱ \times ۰/۰۰۴ = ۰/۴۹۹$$

$$b_2 = b_2 - ۰/۰۱ \times \frac{dE}{db_2} = ۰/۵ - ۰/۰۱ \times ۰/۲۵۷ = ۰/۴۹۷$$

$$w_i = w_i - 0.1 \times \frac{dE}{dw_i} \Longrightarrow$$

$w_1 = 0.1 - 0.01 \times 0.004 = 0.099$

$w_2 = 0.2 - 0.01 \times 0 = 0.2$

$w_3 = 0.3 - 0.01 \times 0.015 = 0.299$

$w_4 = 0.4 - 0.01 \times 0.002 = 0.399$

$w_5 = 0.5 - 0.01 \times 0.019 = 0.498$

$w_6 = 0.6 - 0.01 \times 0.004 = 0.599$

$w_7 = 0.7 - 0.01 \times 0.152 = 0.698$

$w_8 = 0.8 - 0.01 \times 0.225 = 0.797$

$w_9 = 0.9 - 0.01 \times 0.152 = 0.898$

$w_{10} = 0.1 - 0.01 \times 0.247 = 0.097$

new values for weights.