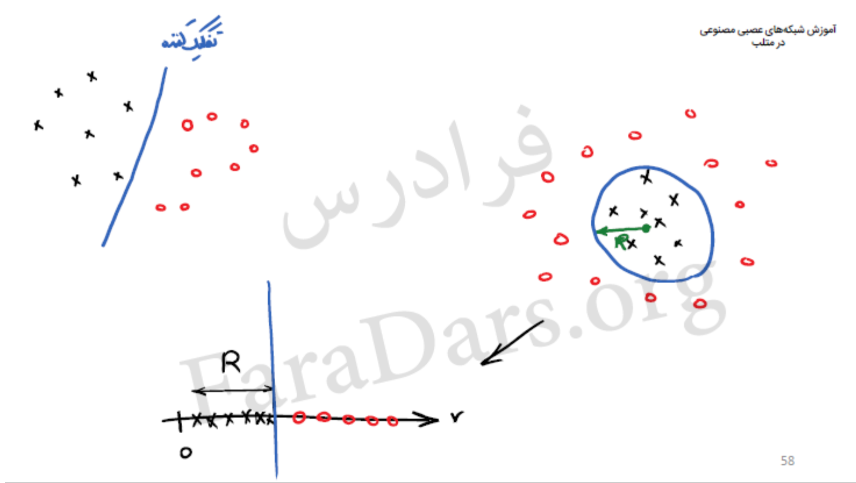
جلسه اول شبکه های عصبی RBF

در اولین صحبت هایی که میتوان از شبکه های عصبی RBF را به جلو بکشانیم این است که هرچیزی که در این دنیا دیده می‌شود پیوسته نیستند ، مثلا همین تصویر تلویزون ،موبایل های ما شامل یک سری پیکسل هایی کنار هم هستند که دارند از خود نور تولید میکنند و از قرار گرفتن کنار یکدیگر میتوان به عنوان یک تصویر دید.

هر کدام از آن ها یک نمایندته از تخمین زننده کلی به حساب میایند .در کل هر کدام از این پیکسل ها نماینده ای به حساب میایند کهع کار خاصی را انجاام دهند یا کواتنتیزه شدن را انجام میدهند.

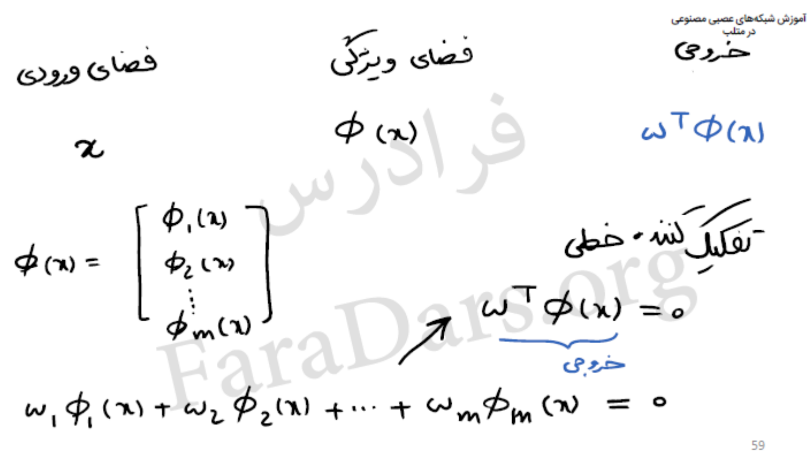
در شبکه های عصبی RBF نیز چنین رویکردی وجود دارد که اطلاعاتت دسته بندی یا کوانتیزه میشوند. یا رویکردی دیگر نگاشت غیر خطی اطلاعات است که اتطلاعات اصلی ما در یک دامنه ای هستند ،سپس با یک نگاشت غیر خطی آن را به یک دامنه ای دیگر امنتقال دهیم .

گاهی اوقات اطلاعات ما به نحوی هستند که برای دسته بندی آنها دچار مشکل میشویم و نمیتوانیم به یک خط به آسانی دسته بندی خود را انحجام دهیم حال برای این موضوع یک رویکردی یا نگاشت غیر خطی نیاز است که باید اطلاعات را از سایر اطلاعات دیگر جداسازی نمود تا بتوانیم عملیات های مختلف روی آنها انجام دهیم .



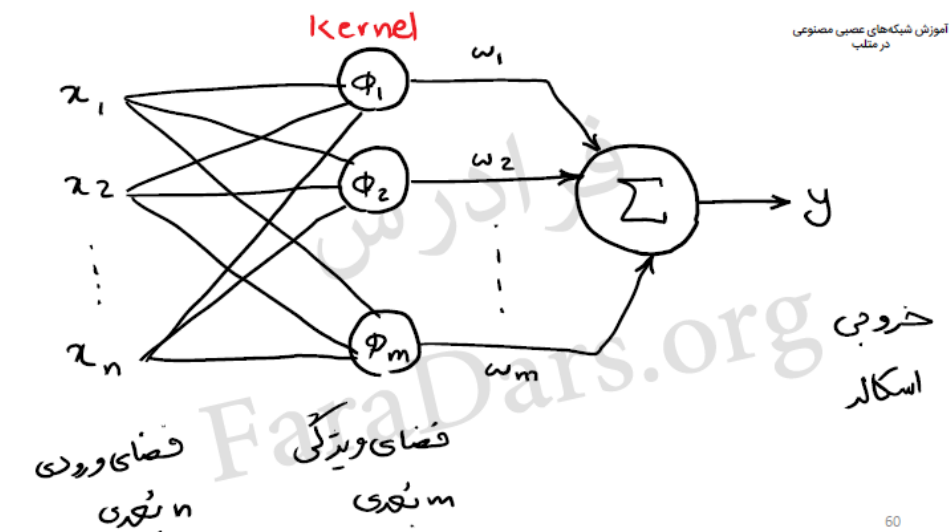
در این تصویر میبنید دسمت راست که اطالاعا دیگر نمیتوان با یک خط آنها را از یکدیگر جداسازی نمود .که این مرز اطلاعات را با r یا شعاع دایره تعین میکنیم. سپس اگر اطلاعات ما کم تر یا مساوی این شعاع دایره باشند را به عنوان دیتای مدنظر میگیریم . در پایین هم یک فضای یک بعدی نمایش داده ایم که بتوانیم کار راحتر شود.

حال اگر بخواهیم به صورت کلی این شبکه عصبی را توضیح دهیم به شکل زیر دقت کنید :



قطعا یک سری ورودی داریم تحت عنوان بردار x سپس این بردار از $\phi(x)$ که این ها یک تر کیب خطی دارند از وزن های خود سپس در هر یک از این $\phi(x)$ ضرب شده و تولید خروجی میکند که در تصویر میبینید.

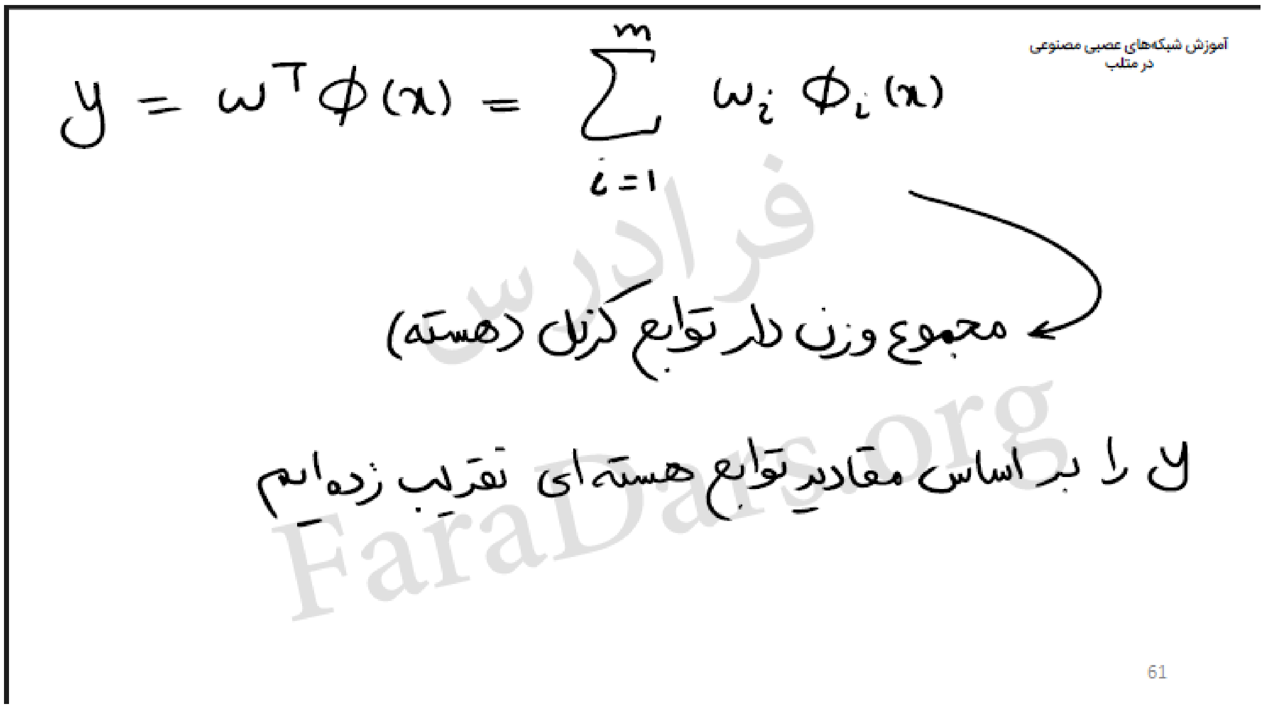
در تصویر بعدی میبینید که این $\phi(x)$ ها هر کدام ویژگی خاصی دارند سپس در خروجی خود در وزن ها ضرب شده در MLP ورودی در وزن ضرب میشد و نرون را تشکیل میداد ولی اینجا کمی متفائت است رویکر RBF :



اینم چیزی هست که برابی شما توضیح دادم .

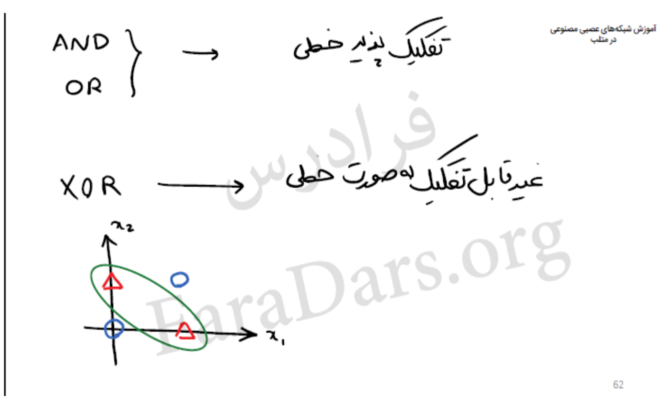
در واقع، این کرنل‌ها عملکردی مشابه توابع فعال‌ساز را برای ما ایفا می‌کنند. نکته مهم اینجاست که تابع $\phi(x)$ چه باشد و آیا خطی یا غیرخطی انتخاب شود؛ این مسئله به نوع مسأله‌ای که با آن سروکار داریم بستگی دارد.

از دیدگاه ریاضی، این فرآیند به معنای تقریب یا بسط خروجی بر اساس مقادیر ویژه‌ی هسته‌ها است که در وزن‌های مربوطه ضرب شده‌اند و در نهایت خروجی را شکل می‌دهند. در شکل زیر میبنید چیزی که را گفتیم :

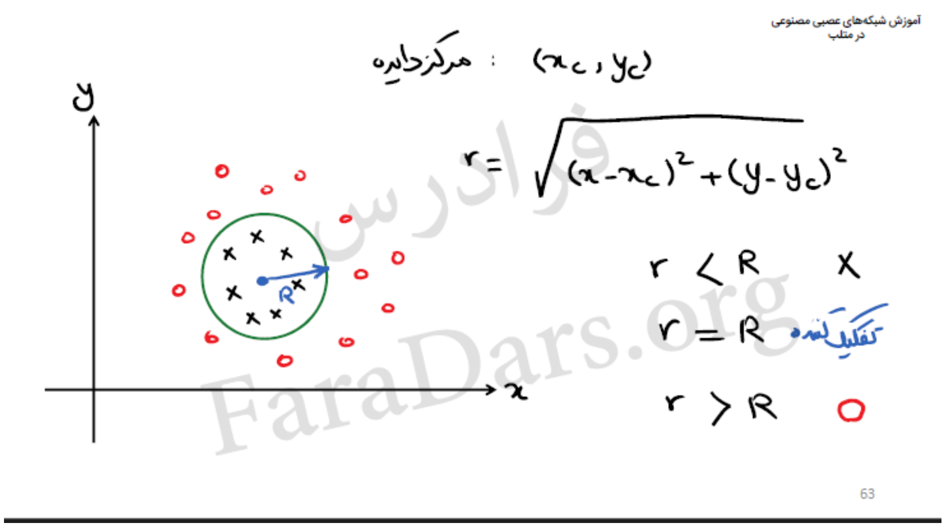


خب در جلسات قبل ما صحبت هایی در رابطه با این که عملگر AND/OR با یک عمگار خطی میتوان دیتاهای خروجی از آنها را دسته بندی کرد اما عملگر XOR چنین رویکردی نمیتواتن داشته و غیرخطی بود به نحوی که اگر میخواستیم با خطی جداسازی کنیم نیازمند به دو خط بود حال در شبکه عصبی RBF میتوان با یک عمگلر غیر خطی آن دیتا ها را دسته بندی کرد و نیاز مند به د و خط راست دیگر نیست .

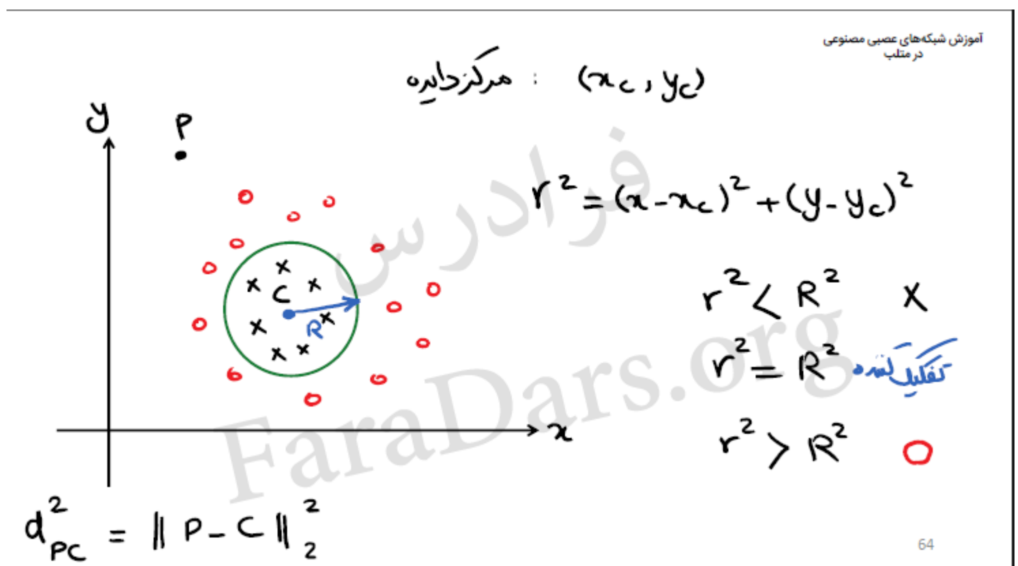
بله در سمت چپ پایین اطلاعات مربوط به خروجی XOR را میبینید که به چه شکلی در آمده است حال با یک دسته بندی بیضی گون که غیر خطی است ما اطلاعات خود را از یکحدیگر جداسازی کرده ایم .



خب در شکل بعدی یک سری اطلاعات را داریم و به صورت RBF میخواهیمک دسته بندی کنیم که به شکل زیر در میاید :

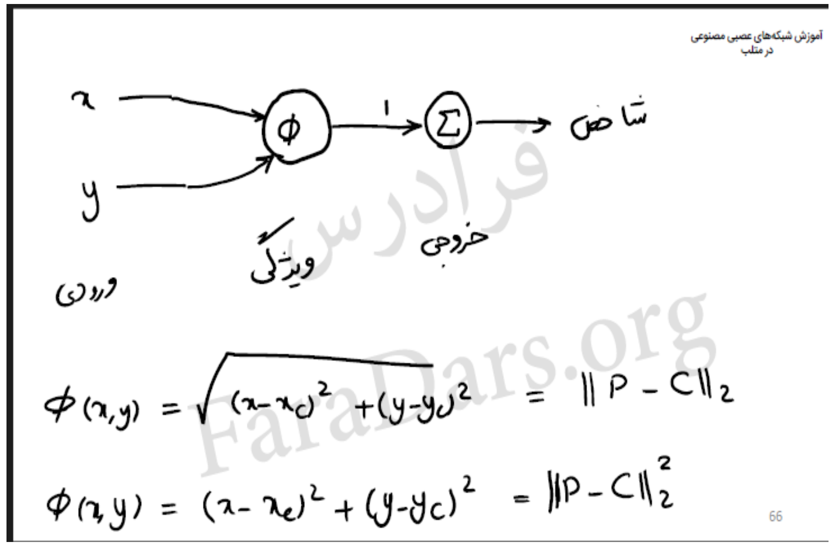


در تصویر مشاهده میکنید که سه ناحیه را براساس شعاع دایره داریم که بیرون داخل و روی مرز د ایر هر چیزی که داخل دایره قرار گیرد جز دسته بندی x هستند و هر چیزی که خارج بیفتد جز O است اما شاید یک اطلاعات پرتی از مرکز دایره داشته باشیم که x باشد حال باید نرم یا فاصله آن را تا مرکز دایره حساب نمود . تصویر گویای همه صحبت های ما است ، به شکل زیر هم میتوان نمایش داد :

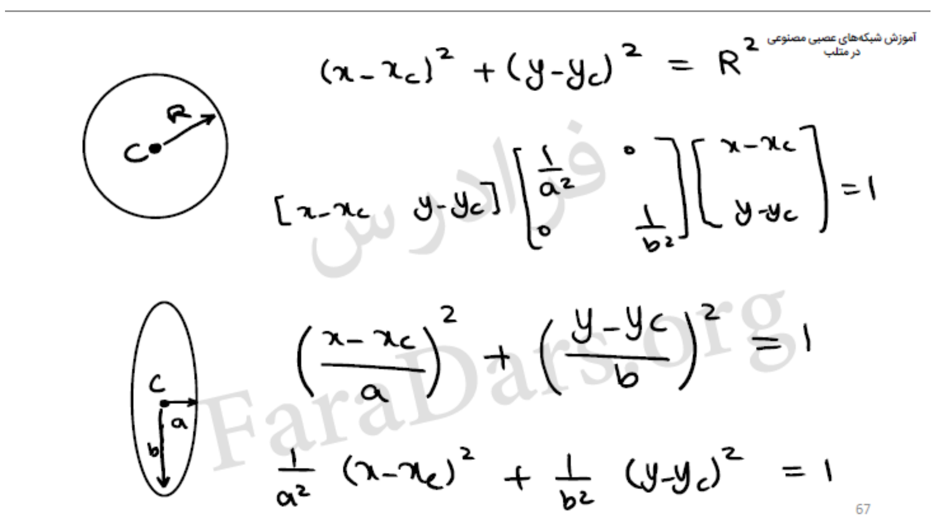


خط در حالت کلی صحبت از $\phi(x)$ همان رابطه درجه یا رادیکالی اتست که برای ما تعین میکنند اطلاعات ما در جا قرار دارند و هستند و در شکل زیر به عنوان یک اسلاید توضیحی بهتر میبینید :

اینم کلیت صحبت که داشتیم که چه اتفاقاتی تابه حال رخ داده است برای سیستم کاری ما .



حال آمدیم اطلاعات به نحوه یا شکل دیگری باشد که نتوان با یک دایره به خوبی آنخها را نمایش داد در حاالت مثلا در شکل زیر اطلاعات به صورت یک بیضی عمودی هستند که باید با یک معادله ای بیان شود این معادله بیضی را به گونه بیان میکنیم که ضرایب مخرج همیشه در معدله بیضی گون حضور دارد به عنوان یک وزن عمل کند حال برای بهتر فهمیدن در شکل زیر توضیحات را آورده ام :



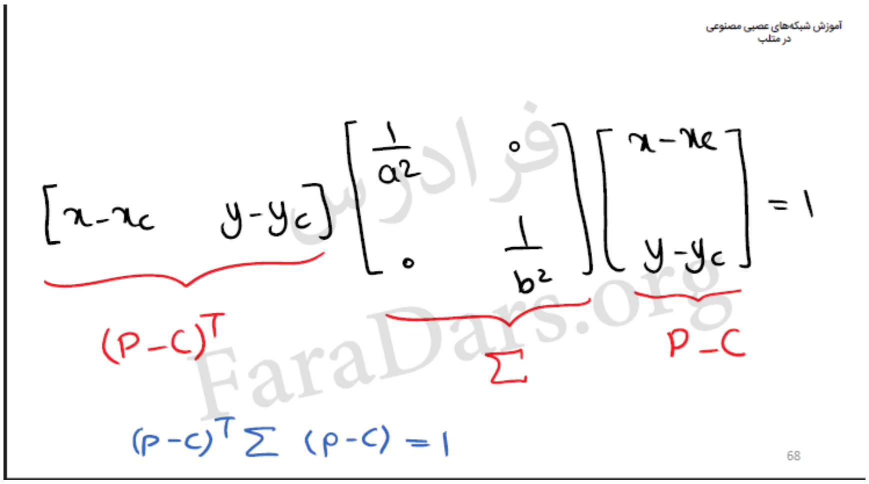
خب این شکل را فکر کنم کفایت کند برای فهمیدن این که میخواهیم چیکار کنیم برای کار های خودمان فقط اونجایی که ماتریس تشکیل داده برای

$$\frac{1}{a^2}, \frac{1}{b^2}$$

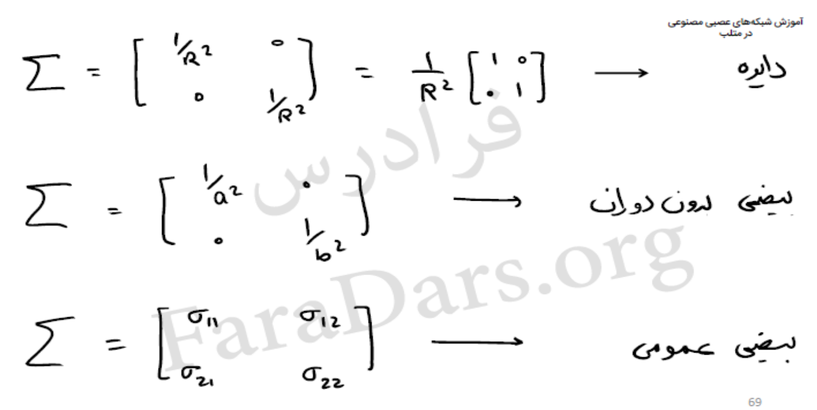
این ماتریس برای این است که کار وزن دهی ما در عملیات های شبکه عصبی راحت تر باشد .

خب شاخصه ما باید یک معادله شود که در تصویر زیر نمایش میدهیم :

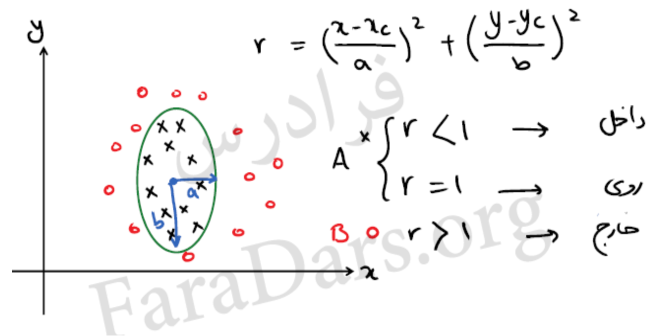
خب در شکل میبنید که به چه نحوی صحبت از شاخص یا یک معادله برای دسته بندی خود کرده ایم در تصویر بعدی نشان میدهیم که اگر هر کدام از این مقادیر ماتریس چقدر باشد چه حالتی به خود میگرد برای دسته بندی

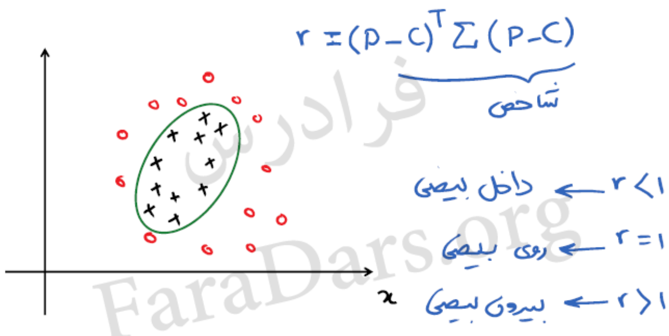


اطلاعات ما :



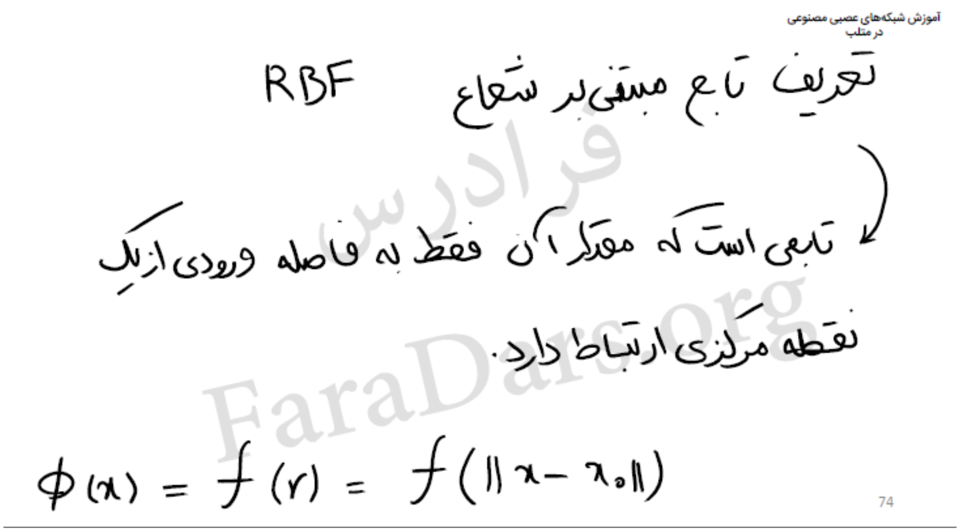
خب د.وتا تصویر میگذارم برای عملکر نوع این شاخص که وقتی که عمودی باشد مثلا دیتاهای x با و زمانی که دیتاهای x ما مایل شوند مثل دو شکل زیر مقایسه کنید که به عهده خودتان است :





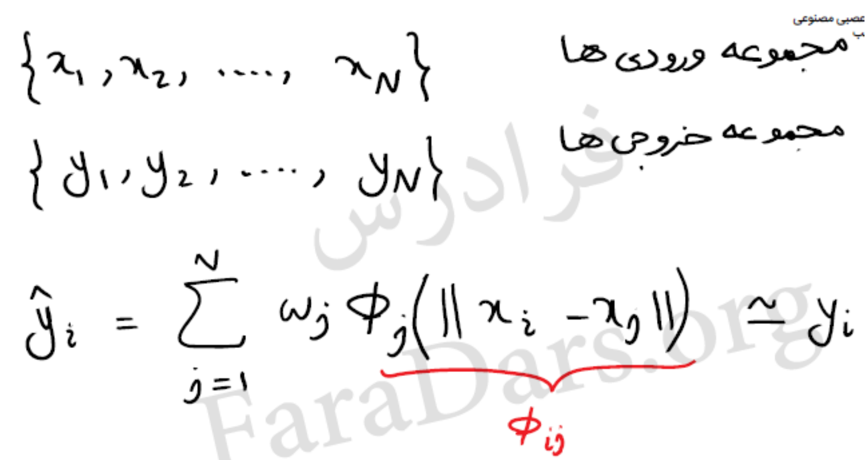
خب در دئو شکل به خوبی فکر کنم دیگر قابل لمس و حس است که چه چیزی رخ دادخه است در آن ماتریس که ه عنوان شاخص کاری ما عمل میکند .

حال یک تعریف کلی از RBF میاوریم که یعنی چی در تصویر زیر آورده ام :



این تقریبا یک تعریف خوبی از شبکه عصبی RBF است شما حتی برای اتم ها و ملکول ها هم میتوان تقریبی از همین فاصله یکه وردی تا یک نقطه مرکزی محاسبه و کار های علمی خود را پیش برد .

در حالت کلی وقتی که با چندین ورودی سروکار داریم در شبکه عصبی RBF دیگر یک معادله جبری سرو کار نداریم بلکه دارای یک دستگاه معادلات است که باید آن را حل نمود که برای توضیحات این بخش برای شما یک سری اسلاید میاورم :



در این اسلاید مجموعه های ورودی و خروجی را نمایش میدهد سپس تابع $\phi(x)$ را باید در نظر گرفت که این تابع مولفه های ماتریس ما میشوند که دارای فاصله اطلاعاتاز مرکز مدنظر است که د ر نهایت به فرم دستگاهی زیر میرسیم :

اینم فرم دستگاه معادلاتی برای چندین ورودی و خروجی است که ماتریس $\phi(x)$ بردار wi ها یلا همان وزن های ما و بردار خروجی ما که این شماتیک کلی یک سیستم RBF است .

$$\begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1N} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{N1} & \phi_{N2} & \cdots & \phi_{NN} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

$$\Phi W = Y \Rightarrow \Phi^{-1}\Phi W = \Phi^{-1} Y$$

$$\Rightarrow W = \Phi^{-1} Y$$

خب حال باید برای برقراری این قضایای فوق میبایست که ماتریس معکوس پذیر باشد که برای این امر یک قضیه هست به اسم $Micchell$ که این قضیه تضمین میدهد که یک سری از توابع ما دارای همین خاصیت معکوس پذیری هستند تابع $\phi(x)$ منظورمان است. در شکل زیر میتوان توابع های گفته شدهخ را دید که بیشتر برای کار های داده ای و شبکه عصبی کاربرد دارند :

شبکه‌های عصبی مصنوعی در متلب

قضیه Micchelli

1) $\phi(r) = \sqrt{r^2 + c^2}$

2) $\phi(r) = \dfrac{1}{\sqrt{r^2 + c^2}}$

3) $\phi(r) = \exp\left(-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2\right)$ گوسی

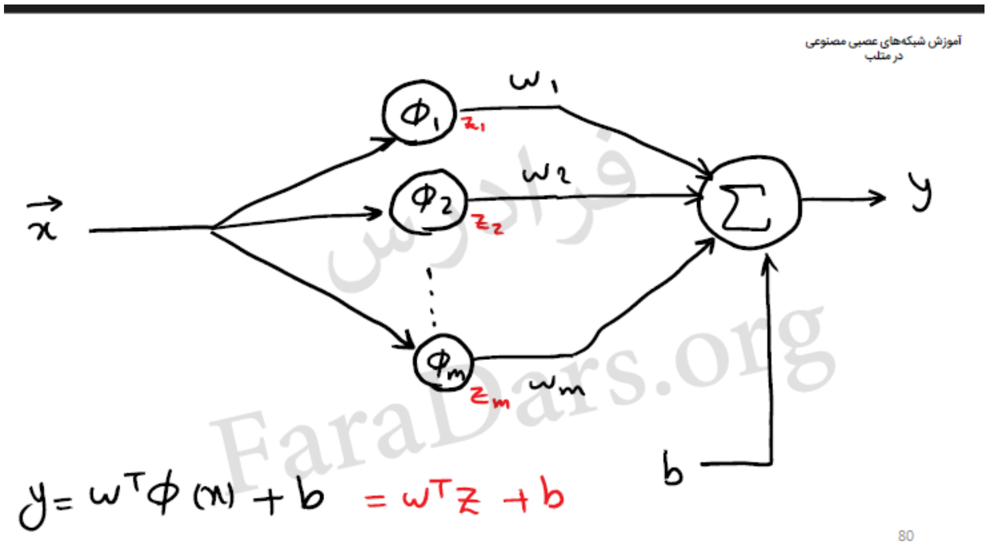
4) $\phi(r) = \dfrac{1}{1 + \exp(-\beta r)}$! $\tanh(r)$ سیگموید

79

اینم از توابعی که میتوان برای این کار انجام آورده ایم .

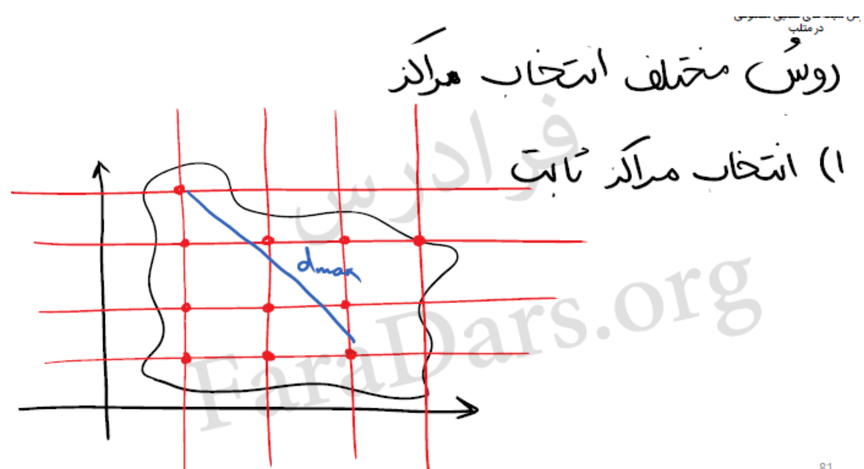
حال دیگر فرم کامل RBF را در قالب یک تصویر به نمایش میگذاریم به شرح زیر است :

اینم تعریفی کلی از این سیستم است .



حال یکی از دردسر های کاری ما در سیستم شبکه عصبی RBF پیدا کردن مرکز است که یک مدل میتوان مرکز را ثابت در نظر گرفت و فرم معادله آن به شکل زیر می‌شود حال اگر تابع $\phi(x)$ ما به فرم گوسی باشد به فرم زیر عمل میکنیم :

در این تصویر مراکز را فیکس یا ثابت در نظر گرفته ایم و در تصویر بعدی نیز میتوان کار کردن برای گاوسی را نشان میدهیم که مرکز داده را کجا انتخاب بکنیم .



$$y = \sum_{i=1}^{m} w_i \phi_i (\|x - x_i\|)$$

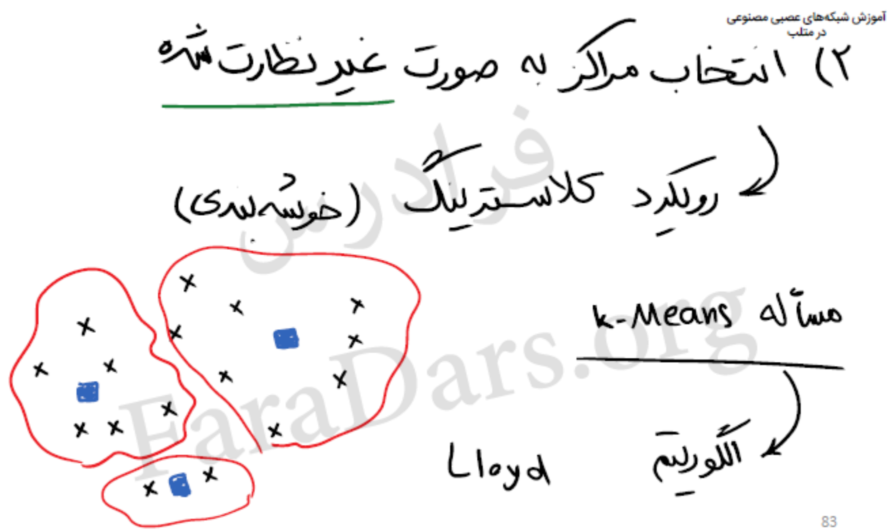
تعداد مراکز

$$\phi_i (\|x - x_i\|) = \exp\left(- \frac{\|x - x_i\|^2}{2\sigma^2}\right)$$

بیشترین فاصله میان دو مرکز

$$\sigma = \frac{d_{max}}{\sqrt{2m}}$$

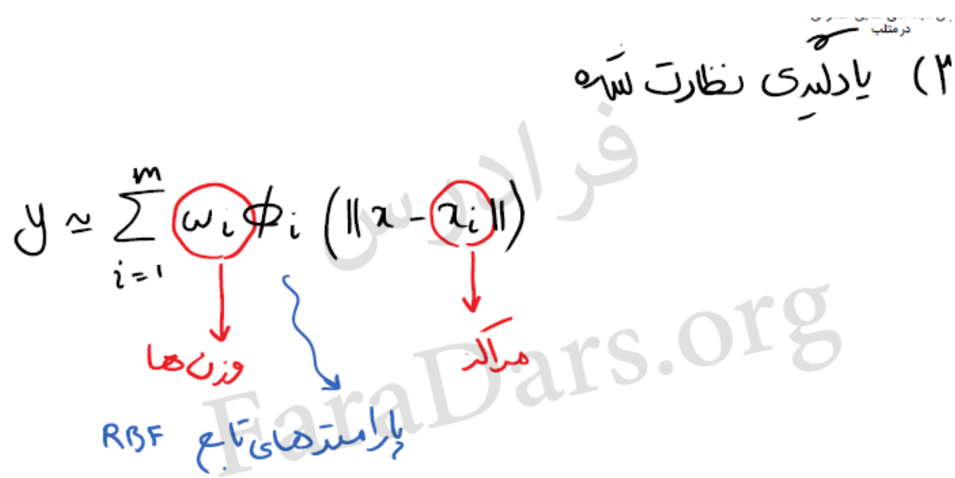
خب یکی از دیگر روش های یکه باید مراکز اطلاعات را پیدا کنیم روش بدون نظارت است که خود الگوریتم بتواند بهترین نقاط را پیدا کند تا بتواند بهترین نتیجه را بر اساس د یتا ها بدهد یکی از این روش های بدون ناظر میتوان به روش *Kmeans* نام برد که در منطق فازی نیز کاربرد داشت همان روشی است که برای *clusturing* داده یا خوشه بندی داده ها کاربرد دارد که این الگوریتم تحت عنوان الگوریتم های بدون ناظر شناخته میشود که در تصویر زیر میبینید :



که در این روش که یکسری نقاط را به عنوان مراکز خود در نظر میگیریمک که این خودش که مسئله بهینه سازی است که آیا این مرکز به عنوان یک جای مناسب برای مرکزیت داده است یا خیر سپس بر اساس آن دیتا ها خوشه بنید میشوند تا شعاعی بازهم بر اساس بهینه سازی شده .

خب در حالت کلی یادگیری نظارت شده به شکل زیر است :

این یک سیستم یادگیری با نظارت *RBF* میباشد.



خب برای این که ببینیم این شبکه عصبی خوب کارکرده اسیا خیر باید بیاییم ببینیم که خطای ما بین ما و سیستم اصلی چقدر است و حال بعد از این باید خطای خود و سیستم را حداقل کرد که این دوباره یک مسئله بهینه سازی است .

پارمتر های قابل د ستکاری ما یکی وزن های شبکه عصبی هستند و دوم مراکز دیتاهای ما میباشد که باید تنظیم کاری کردن این دو پارامتر میتوان مسئله بهینه سازی را برای خطای شبکه عصبی در نظر گرفت .

آموزش شبکه‌های عصبی مصنوعی در متلب

$$y_i \simeq \hat{y}_i = \sum_{j=1}^{m} w_j \phi_j(\|x_i - x_j\|)$$

وزن‌ها　　مرکزها

$$e_i = y_i - \hat{y}_i$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N} e_i^2 \longrightarrow \text{Performance Index}$$

85

حال فرض کنید دیتاهای ما تابع گوسی هستند و میخواهیم الگوریتم RBF را برای این دیتاها بزنیم ،در این حالت باید چیکار کنیم خب برای این کار ابتدای امر وزن آن را در زمان صفرم پیدا باید بکنیم سپس باید مسئله خطا یا در نظر بگیریم که حداقل شود سپس برای یکی زمان جلوتر خود باید خود زمان صفرم با یک ضریبی در مشتق خطا ضرب شود کم کنیم تا وزن زمان بعدی را بدست بیاورد یم که در شکل زیر آورده ام :

آموزش شبکه‌های عصبی مصنوعی در متلب

$$\phi_j(r) = \exp\left(-\frac{r^2}{2\sigma^2}\right)$$ گوسی

$$y_i \simeq \hat{y}_i = \sum_{j=1}^{m} w_j \phi_j(\|x_i - x_j\|)$$

$$SSE = \sum_{i=1}^{N} e_i^2 \qquad w_j(0)$$

$$w_j(t) \longrightarrow w_j(t+1)$$

$$w_j(t+1) = w_j(t) - \eta \left.\frac{\partial SSE}{\partial w_j}\right|_{w_j(t)}$$

86

این تمام چیزی هست که در بالا به صورت متنی آورده ام .

حال مشتق خطا را به فرم زیر تعریف میکنیم :

$$\frac{\partial SSE}{\partial w_j} = \frac{\partial}{\partial w_j}\left[\sum_{i=1}^{N} e_i^2\right] = \sum_{i=1}^{N} \frac{\partial e_i^2}{\partial w_j}$$

$w_j(0) \longrightarrow$ مقدار اولیه

$$w_j(t+1) = w_j(t) - \eta_j \frac{\partial SSE}{\partial w_j}$$

G.D. روش گرادیان نزولی

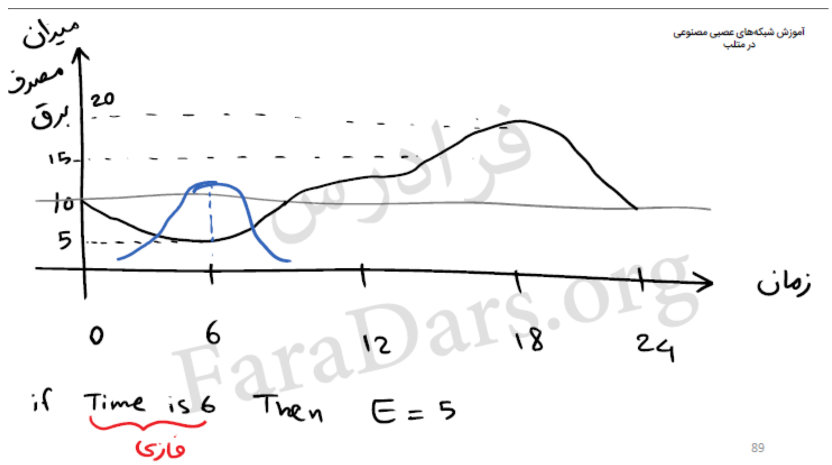
خب این الگوریتم همان روش گرادیان نزولی است که یک سری نقاط مینیمم محلی و کلی را میدهد که ما به دنبال مینیمم های کلی هستیم در حالت اول حال اگر شرط اولیه درست دهیم دوم این که نقطه ای که بدست میاید ناحیه جذب دارد که نقطات را در آن ناحیه مینیم به سمت خودش میکشد که باید کاری کنیم به سمتی جذب شود این نواحی جذب که به سمت ناحیه مینیم کلی برویم .

خب یکی از رویکرد های دیگری که میتوان برای شبکه های عصبی *RBF* در نظر داشت همان رویکرد منطق فازی است که در ریپازیتوری *fuzzy logic* من کاملا با آن بحث و سخن ها کار کرده ایم ولی اینجا یکی از چیز های مهم که از منطق فازی داریم این است که یک سری قوانین را بر حسب فازی بیان میشودئ که *mamdani* نام دارد که هما اگر و انگاه خودمان است ولی یک منطق فازی دیگری داریم که بر حسب توابع ریاضی و رویکرد ریاضی است *TSK* یا سگنو میباشد .

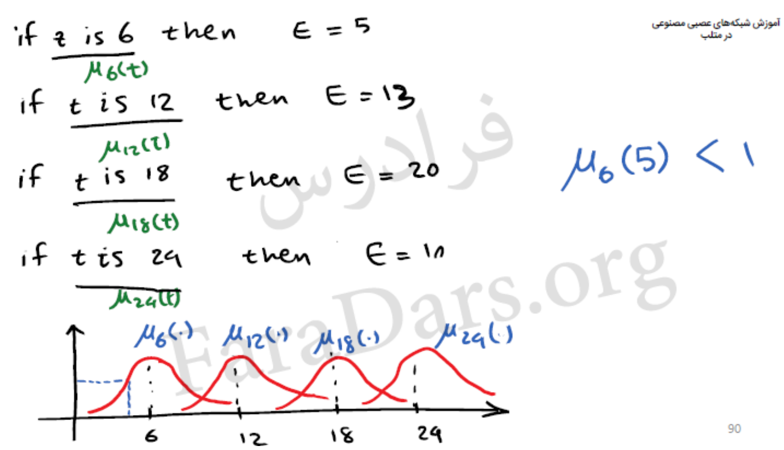
که برای این بخش دیگر تئوری کار نمیاورم که اگر بیاید داشته باشید ما یک چیزی تحت عنوان *membership function* یا درجه تعلق داشتیم که همان ها به عنوان ضرایب وزنی در نظر میگیریم و میدانستیم برای هر کدام از $\mu(x)$ یک نتیجه دارد که آن نتیجه را به عنوان خروجی کار در نظر میگیریم و سپس خروجی را در هر کدام از این $\mu(x)$ ضرب وبر کل مجموع $\mu(x)$ تقسیم میکنیم .

مثلا ما برای اطلاعات یک مصرف نیرو. در یک کشور اطلاعات جمع آوری کرده ایم سپس در بازه های ساعتی یا همان زمانی مختلف میزان مصرف را در میاوریم و نکته ای که داریم *step* های زمانی ما مشخص است مثلا 6 تایی باشد یا هر چند عددی که است ولی تقسیم بندی منطم باشد سپس مقادیر را میخوانیم که بر حسب ساعت چثدر مصرف نیرو داریم حال در زمان های بعدی هم همین کار را میکنیم و میتوانیمک میزان مصرف را به صورت یک تابع گوسی از مصرف نیرو و زمان بیرون کشید در هر *step* زمانی ما که برای این به شکل زیر مراجعه کنید :

خب این نمودار ها را برا یزمان های دیگر هم رسم نموده و سپس شکل زیر را در میاوریم :



میزانن مصرف در زمان های مختلف را بر حسب تابع گوسی به صورت فازی بیان شده است .حال این *mu (x)* های ما به عنوان ضرایب و خروجی ها هم *E* است که به فرمول زیر میخواهیم کلیت کار را در بیاوریم :



حال اگر بتونیم مجموع *membership functions* ها به صورتی تعرف کنیم که یک شود به صورت خیلی عالی در میاتید که سیستم ما خوب عمل کند.