تمرین سری چهارم درس داده کاوی - فاطمه غلام‌زاده - ۹۵۳۱۰۶۰

سوال ۱

قضیه همگرایی: اگر مجموعه وزن‌های w^* ها جدایی پذیر خطی باشند (یعنی قابلیت جداسازی یک مجموعه‌ی خطی را داشته باشند)، الگوریتم یادگیری پرسپترون همگرا خواهد شد.

قضیه: اگر بردار وزن w^* را به گونه‌ای داشته باشیم به ازای هر p:

$$f(x(p) \cdot w^*) = t(p)$$

آن گاه برای هر بردار اولیه w الگوریتم یادگیری پرسپترون به بردار وزن تردیک می‌شود که برای همه‌ی الگوها پاسخ صحیحی می‌دهد (در مراحل متناهی)

$t(p)$: مقدار هدف معادل بردارها ورودی آموزش

p: تعداد بردارهای ورودی

$x(p)$: بردارهای ورودی

f: حدودی activation function

نکته: وزن‌هایی که الگوریتم به آن‌ها نزدیک می‌شود لزوما منحصر به فرد و لزوما برابر با w^* نیستند.

سوال ۲

$$\text{prediction}(y') = \begin{cases} 1 & wx+b \geq 0 \\ 0 & wx+b < 0 \end{cases}$$

گیت NOR فقط در صورتی که هردو ورودی ۰ باشند ۱ می‌دهد.

NOR: 

A	B	$\overline{A+B}$
0	0	1
0	1	0
1	0	0
1	1	0

سطر اول:

مقدار اولیه: $w_1 = 1$, $w_2 = 1$, $b = -1$

Model: $x_1(1) + x_2(1) - 1$

سطر اول: $x_1 = 0$, $x_2 = 0$ $\Rightarrow 0+0-1 = -1 \Rightarrow wx+b = -1 < 0 \Rightarrow y' = 0$

با توجه به خروجی جدول کاری خواهیم داشت غلط شد ✗ پس b را تغییر

می دهیم و دوباره از دستور کار می کنیم: $0+0+1=1>0 \Rightarrow y'=1$ ✓

سطر دوم:

$x_1=0 \quad x_2=1 \Rightarrow wx+b=0+1+1=2>0 \Rightarrow y'=1$

↳ طبق ردیف دوم خروجی باید صفر می شد پس غلط ✗ این بار w1 و w2 را

۲ - ۱ - تغییر می دهیم و $b=0/5$

$wx+b=0-1+0/5=-0/5<0$

↳ $y'=0$ ✓

سطر سوم:

$x_1=1 \quad x_2=0$

$wx+b=-1+0+0/5=-0/5<0 \Rightarrow y'=0$ ✓

با توجه به ردیف سوم که صفر است جواب درستی حاصل شد

سطر چهارم:

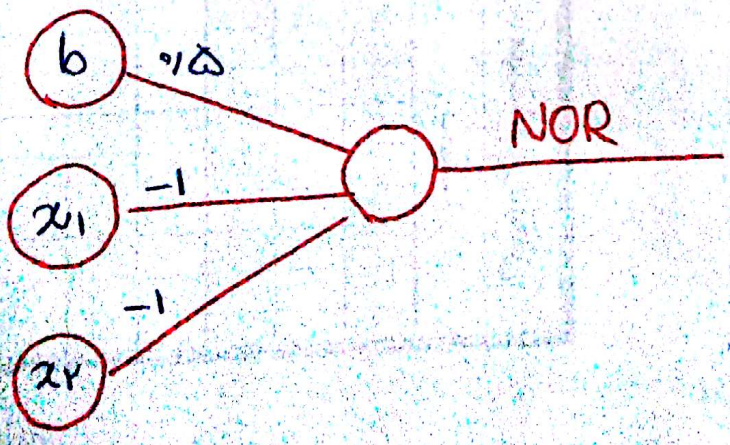
$x_1=1 \quad x_2=1$

$wx+b=-1-1+0/5=-1/5<0 \rightarrow y'=0$ ✓

NOR Gate Model:

$$w_1=w_2=-1$$
$$b=0/5$$

$$y'=\begin{cases}1 & -x_1-x_2+0/5\geq 0\\ 0 & -x_1-x_2+0/5<0\end{cases}$$



NAND:

A	B	out
0	0	1
0	1	1
1	0	1
1	1	0

سطر اول:

مقداردهی اولیه‌ی w_1, w_2, b:

$$\begin{cases} w_1 = w_2 = 1 \\ b = -1 \end{cases}$$

$x_1 = x_2 = 0 \rightarrow wx+b = 0+0+(-1) = (-1) < 0 \rightarrow y' = 0$ ← غلط

b را آپدیت می‌کنیم ← $b = 1$:

$wx+b = 0+0+1 = 1 > 0 \rightarrow y' = 1$ ✓

سطر دوم:

$x_1 = 0 \quad x_2 = 1 \rightarrow wx+b = 0+1+1 = 2 > 0 \rightarrow y' = 1$ ✓

سطر سوم:

$x_1 = 1 \quad x_2 = 0 \Rightarrow wx+b = 1+0+1 = 2 > 0 \rightarrow y' = 1$ ✓

سطر چهارم:

$x_1 = x_2 = 1$

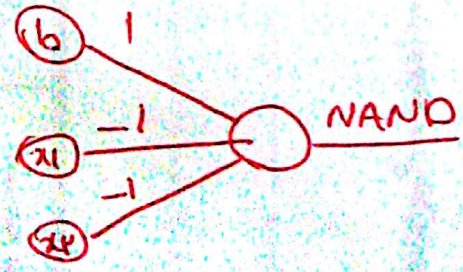
$wx+b = 1+1+1 = 3 > 0 \rightarrow y' = 1$ X

w_1 و w_2 را آپدیت می‌کنیم

$\boxed{w_1 = w_2 = -1}$

$wx+b = -1-1+1 = -1 < 0 \rightarrow y' = 0$ ✓

NAND Gate Model:



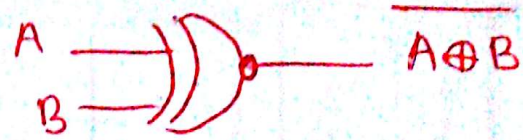
$w_1 = w_2 = -1$

$b = 1$

$$y' = \begin{cases} 1 & -x_1 - x_2 + 1 \geq 0 \\ 0 & -x_1 - x_2 + 1 < 0 \end{cases}$$

XNOR :

A	B	out
0	0	1
0	1	0
1	0	0
1	1	1



گیت xnor از گیت های AND و NOR و OR تشکیل شده :

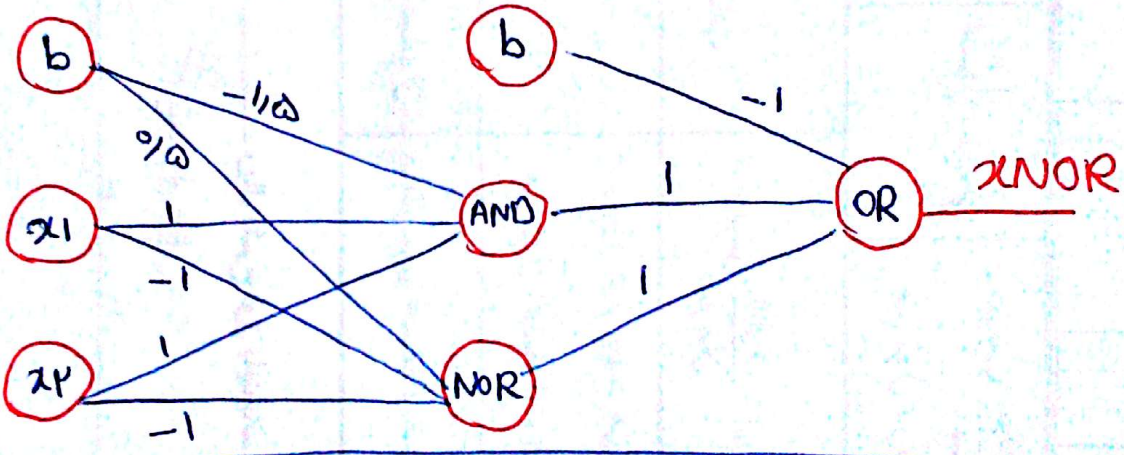
پس باید پرسپترون های آن ها را با هم ترکیب کنیم.

$$\overline{A \oplus B} = \underbrace{AB}_{AND} + \underbrace{A'B'}_{NOR} \quad \text{(NOR)}$$

AND $\rightarrow x_1 + x_2 - 1.5$

NOR $\rightarrow -x_1 - x_2 + 0.5$

OR $\rightarrow x_1 + x_2 - 1$



سوال ۳ الگوریتم Kmean++ : تفاوت این الگوریتم با K-means

فقط در انتخاب مختصات اولیه مراکز خوشه ها است. ابتدا یکی از داده ها به طور تصادفی انتخاب شده و به عنوان مرکز یک خوشه در نظر گرفته می شود. سپس مراکز بعدی با احتمالی که با مربع فاصله آن داده با نزدیکترین مرکز خوشه از بین مراکزی که تا به حال انتخاب شده اند، متناسب است، انتخاب می شود و این روند ادامه می یابد تا k تا مرکز انتخاب شود. این الگوریتم در ابتدا زمان بیشتری را صرف انتخاب مراکز می کند (نسبت به kmean) اما این کار بعدا باعث می شود الگوریتم زودتر همگرا شود و همیشه سرعت بیشتری از K-means دارد و خطای کمتری هم دارد. زمان اجرا: $O(\log(k))$

سوال ۴ در ابعاد بالا ۲ روش برای بصری سازی اطلاعات داریم :

① PCA ② T_SNE

① PCA ← در این روش دیتا را به فضای دیگری می بریم، component ها را براساس اهمیتشان Sort کرده و component های اضافی را حذف می کنیم به گونه ای که کمترین بخش عمده ای از دیتای اصلی توسط این component های اصلی قابل بازیابی است. این روش برای بصری سازی کلاسترهای فضای اقلیدسی ، ۲ و ۳ بعدی به ما کمک می کند. این روش بیشتر برای اطلاعات تا ۳ بعد مناسب است و در ابعاد بالا خوب عمل نمی کند.

② T_SNE : برای داده های با ابعاد بالا مناسب است. از ارتباط همسایگی بین نقاط برای نگاشت به ابعاد پایین تر استفاده می کند و ساختاری غیرخطی دارد. این روش یک توزیع احتمال (مثلا گوسی) ایجاد می کند که نشان دهنده ی ارتباط میان نقاط همسایه است ، سپس یک فضای با ابعاد کم را می سازد که آن توزیع را تا جایی که می تواند حفظ کند و از ست t استفاده می کند.

یک روش دیگر برای نمایش اطلاعات در ابعاد بالا ، Correlation heatmap نام دارد. با رسم کردن هیستوگرام برای هر فیچری توان ایده گرفت که هر بعد چگونه توزیع شده است.

Correlation heatmap نشان دهنده ی چگونگی ارتباط دو بعد از داده ها باهم است. اگر چند هیستوگرام ۲ بعدی زیاد باشد نشان می دهد که ویژگی ها بهم وابسته اند و اگر کم باشد استقلال بین آن ها را نتیجه می دهد.

برای نشان دادن داده‌ها در ۳ بعد برای ویژگی سوم می‌توانیم از شکل، رنگ و ... استفاده کنیم.