



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
دانشکده مهندسی کامپیوتر و فناوری اطلاعات

تمرین اول

مبانی هوش محاسباتی

نگارش

محمدرضا اخگری زیری

استاد درس

دکتر عبادزاده

فروردین ۹۹

صفحه

فهرست مطالب

سوال اول.....	۳
سوال دوم.....	۵
سوال سوم.....	۸
سوال چهارم.....	۱۱

سوال اول

هر نورون از سه قسمت اصلی تشکیل شده است:

۱. بدنه سلول (Soma)

۲. دندريت (Dendrite)

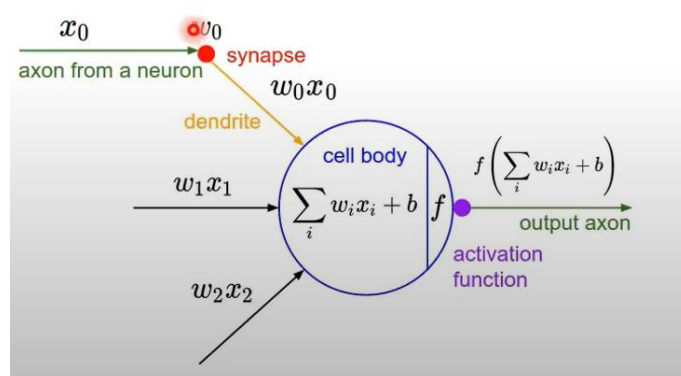
۳. اکسون (Axon)

دندريت‌ها به عنوان مناطق دريافت سيگنال‌های الکتریکي، شبکه‌هایی تشکیل یافته از فیبرهای سلولي هستند که دارای شاخه‌های انشعابی بیشماری هستند. دندريت‌ها **سيگنال‌های الکتریکي** را به هسته سلول منتقل می‌کنند.

بدنه سلول انرژی لازم را برای فعالیت نورون فراهم می‌کند. بر روی سيگنال‌های دريافتي عمل (یک عمل ساده جمع و مقایسه مدل می‌شود) می‌کند.

اکسون طول بیشتری دارد و **سيگنال‌های الکتروشیمیایی** دريافتي از هسته را به نورون‌های دیگر منتقل می‌کند.

سیناپس محل تلاقی یک اکسون یک سلول به دندريت‌های سلول‌های دیگر است و ازتباطات مابین نورون‌ها را برقرار می‌کند. این فاصله برای این است که سيگنال خروجی را تقویت کنند.



در مدل سازی نورون مصنوعی، می توان ناحیه سیناپسی را با وزن های داده شده به مقادیر ورودی (W) نشان داد.

ورودی ها به نورون برابر با دندریت ها هستند.

و بدنه سلول همان معادله ریاضی است.

و خروجی حاصل از تابع فعالیت، خروجی ای است که آکسون باید تولید کند.

به شکل صفحه قبل مراجعه کنید.

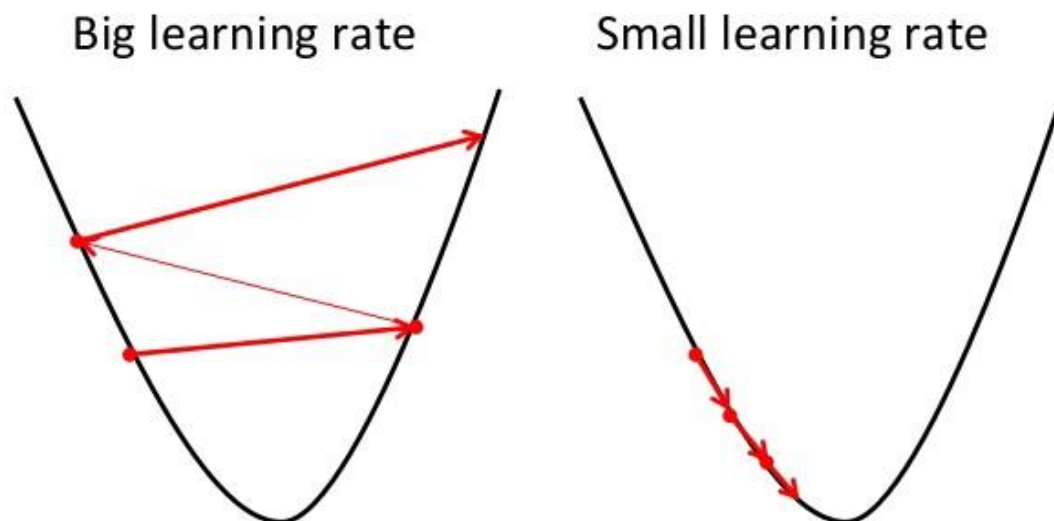
سوال دوم

(الف)

در الگوریتم گرادیان کاهشی برای یافتن وزن‌های مناسب ($w^{\text{new}} = w^{\text{old}} - \tau \nabla E(wt)$) نرخ یادگیری، اندازه‌ی گام‌هایی است که برای پیدا کردن وزن جدید به کار می‌رود، اگر این مقدار کوچک باشد الگوریتم زمان بسیاری طول می‌کشد و اگر بزرگ باشد، امکان دارد در نزدیک نقاط مینیمم محلی الگوریتم در حلقه بینهایت بیفتد، پس میتوان در مناطق دور از مینیمم محلی مقدار این عدد را بزرگ گرفت و در نزدیکی مینیمم محلی مقدارش را کوچک در نظر گرفت.

میتوان شکل زیر را در نظر گرفت.

Gradient Descent



(ب)

افزایش تعداد لایه های پنهان ممکن است دقت را بهبود بخشد یا ممکن است نباشد ، واقعاً به پیچیدگی مسئله ای که می خواهیم حل کنید بستگی دارد. افزایش تعداد لایه های پنهان بسیار بیشتر از تعداد کافی لایه ها باعث می شود دقت در مجموعه تست کاهش یابد. یا بیش‌برازش اتفاق بیفتد.

(ج)

با افزایش عرض یک لایه، اطلاعاتی می‌دهیم که اگر این اطلاعات زیاد باشد باز هم باعث بیش‌برازش می‌شود و مثل مورد قبلی باعث افزایش زمان پردازش و هزینه زیاد می‌شود.

(د)

راه‌های فهمیدن بیش‌برازش:

• اعتبارسنجی متقاطع^۱

برای تشخیص بیش‌برازش در مدل می‌توان از «اعتبارسنجی متقاطع» استفاده کرد. در این حالت داده‌های نمونه به دو یا چند بخش تفکیک شده و در هر مرحله یکی از بخش‌ها برای برآورد پارامترهای مدل به کار می‌رود. این بخش از نمونه را «مجموعه داده‌های آموزشی (Training Set)» می‌نامند.

بخش‌های دیگر نمونه که به آن «مجموعه داده‌های آزمایشی (Test Set)» می‌گویند برای سنجش میزان خطای پیش‌بینی مدل به کار می‌روند. روند اعتبارسنجی متقاطع به صورت زیر است:

۱. داده‌ها را به دو بخش آموزشی و آزمایشی تفکیک می‌کنیم.
۲. برای داده‌های آموزشی پارامترهای مدل مناسب را براساس کمینه سازی تابع خطا، برآورد می‌کنیم.
۳. خطای برازش مدل ایجاد شده را روی داده‌های آزمایشی اندازه‌گیری می‌کنیم.
۴. نسبت خطای بدست آمده از مدل، برای داده‌های آزمایشی و آموزشی نباید خیلی بزرگ باشد.

^۱ Cross-Validation

- مراحل ۱ تا ۴ را با توجه به همگرا شدن نسبت حاصل از مرحله ۴ ادامه می‌دهیم در غیر این صورت به تعداد تکرار مشخص، عملیات پایان می‌یابد (هر کدام زودتر به وقوع بپیوندند). با توجه به میزان نسبت خطاهای ذکر شده بهترین مدل در این مرحله حاصل می‌شود.

روش جلوگیری از بیش برآزش:

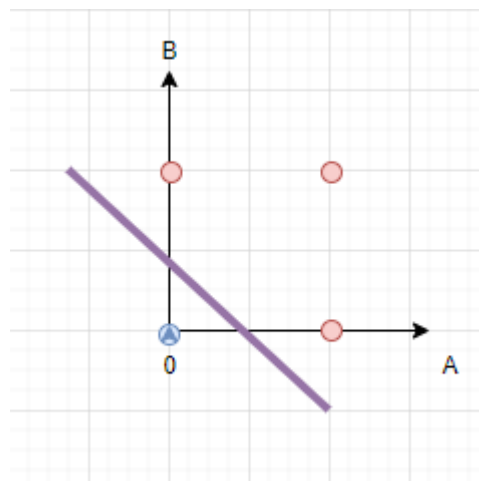
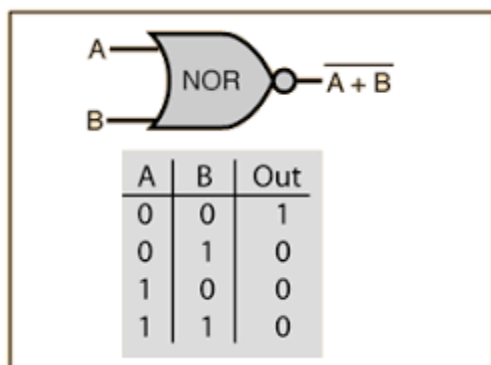
حذف تصادفی به تکنیکی اطلاق می‌شود که برخی گره‌های (node) شبکه را به صورت تصادفی حذف می‌کند. حذف تصادفی را می‌توان به صورت غیرفعال سازی موقت یا نادیده گیری نوروں‌های شبکه در نظر گرفت. این تکنیک در فاز آموزش، جهت کاهش بیش برآزش بکار می‌رود.

این طور فرض کنید که هر روز با تعداد زیادی از افراد ملاقات می‌کنید. درحالی که شخصا با آن‌ها صحبت می‌کنید، چهره را به خاطر می‌سپرید. گاهی اوقات ارتباط شما به اجبار تلفنی است. همان افراد را این بار تشخیص نمی‌دهید، چون آن‌ها را فقط دیده‌اید و چهره‌شان را به یاد دارید. شما، در ارتباط تلفنی صدایشان را به خاطر می‌سپارید و به یاد می‌آورید. بنابراین، شما، با حذف تصادفی ویژگی‌های بصری، مجبور می‌شوید روی ویژگی‌های صوتی تمرکز نمایید.

عوض کردن ضریب جمله تنظیم، یکی دیگر از روش‌هاست برای جلوگیری از بیش‌برآزش. در روش گرادینان کاهشی می‌توان جمله تنظیمی نیز بگذاریم که هر چقدر ضریب آن بیشتر باشد، به جمله‌های خارج از مجموعه آموزشی بیشتر اهمیت می‌دهد. (برای مواقعی که نم‌خواهیم نوروں‌ها را حذف کنیم)

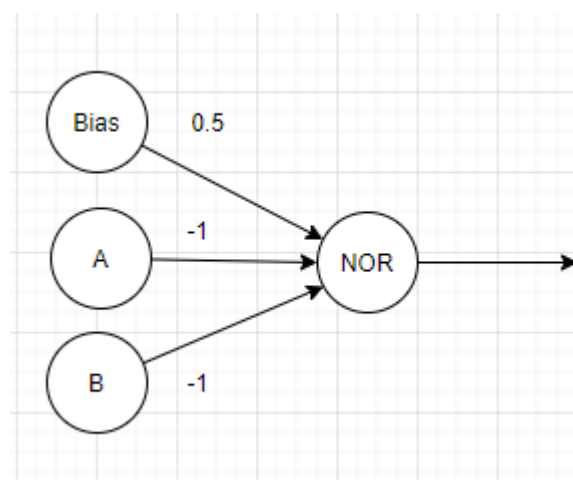
سوال سوم

NOR:

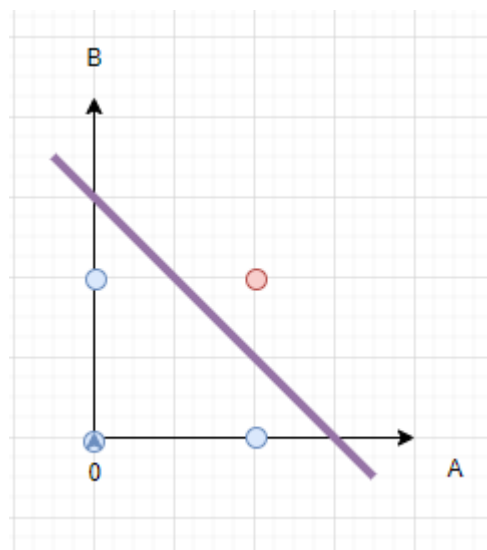
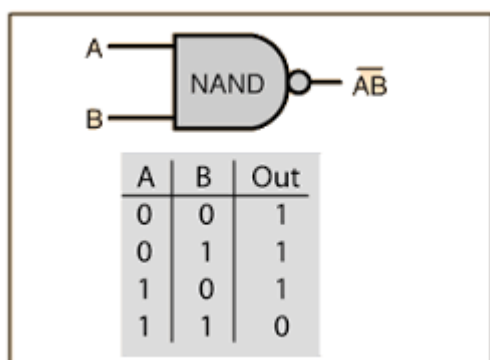


$$-A - B + 0.5 \leq 0$$

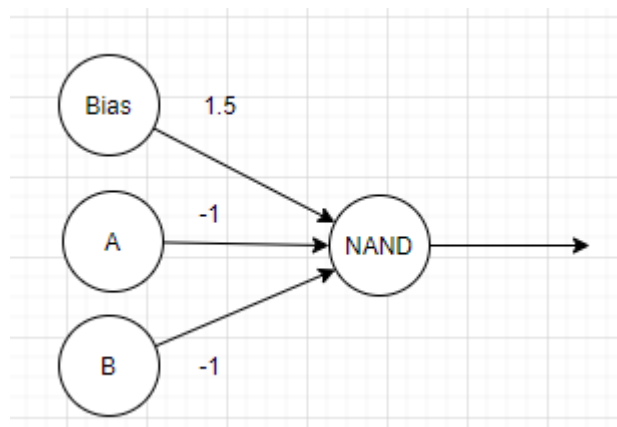
البته میشد که خط را از نقاط وسط رد کرد (بایاس را یک گرفت)



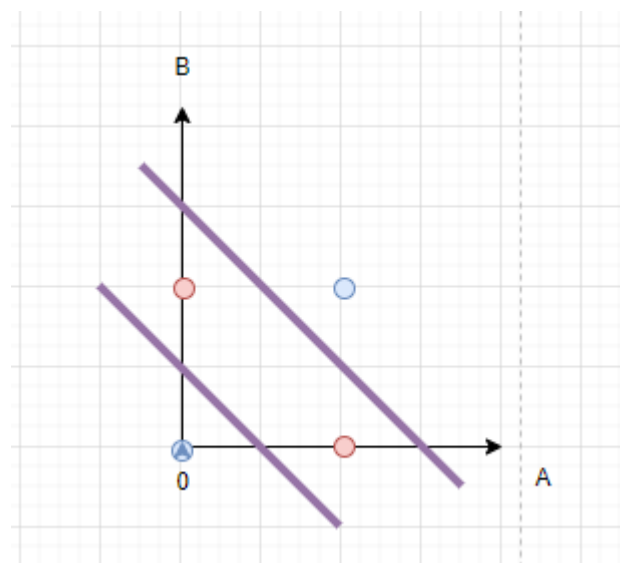
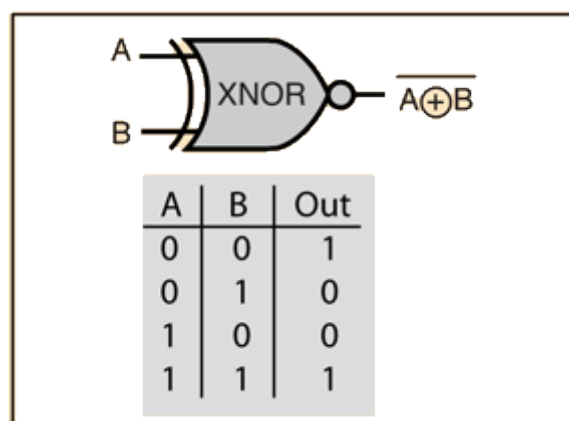
NAND:



$$-A - B + 1.5$$



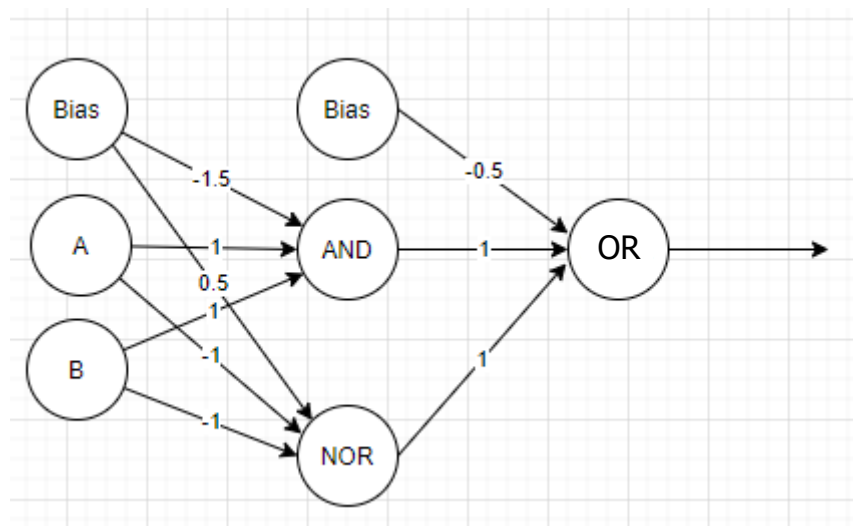
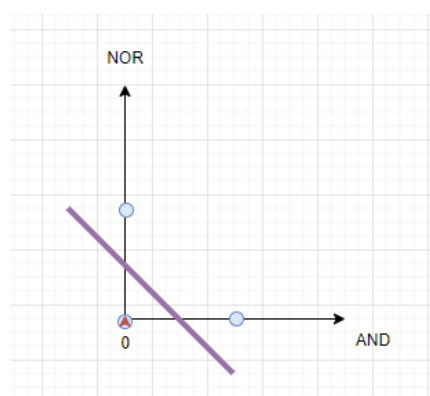
XNOR:



میتوان فهمید که با یک لایه شبکه عصبی نمی‌توان مدل سازی کرد، لذا به از دو لایه استفاده میکنیم و میدانیم که:

$$XNOR = (A \text{ and } B) \text{ or } (A \text{ nor } B)$$

می‌توان از نمودار and و nor استفاده کرد.



سوال چهارم

یادگیری بدون ناظر:

وزن‌ها بر اساس ورودی اصلاح می‌شوند و خروجی مطلوب وجود ندارد تا با مقایسه خروجی شبکه با آن و تعیین مقدار خطا وزن‌ها اصلاح شود. هدف استخراج مشخصه‌های الگوهای ورودی براساس استراتژی خوشه‌یابی و یا کلاس‌بندی و تشخیص شباهت (تشکیل گروه‌هایی با الگوهای مشابه) می‌باشد، بدون اینکه خروجی یا کلاس‌های متناظر با الگوهای ورودی از قبل مشخص باشد.

یادگیری باناظر:

به ازای هر دسته از الگوهای ورودی خروجی‌های متناظر نیز به شبکه نشان داده می‌شود و تغییر وزن‌ها تا زمانی انجام می‌شود که اختلاف خروجی از خروجی موردنظر خطای قابل‌قبولی داشته باشد. چنین شبکه‌هایی بطور گسترده برای کارهای تشخیص الگو به کار گرفته می‌شود.

یادگیری تقویتی:

کیفیت عملکرد سیستم به صورت گام به گام نسبت به زمان بهبود می‌یابد. الگوهای آموزشی وجود ندارد اما با استفاده از سیگنالی به نام نقاد بیانی از خوب و یا بد بودن رفتار سیستم بدست می‌آید (حالتی بین یادگیری با سرپرست و بدون سرپرست). می‌توانیم آن را یادگیری صحیح و خطا هم بنامیم، چرا که سیستم عملی را انجام می‌دهد و در صورتی که آن عمل درست باشد، آن را تکرار خواهد کرد و در غیر اینصورت تلاش می‌کند تا عمل جدیدی به دست آورد.

