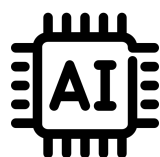


بسم تعالی



# هوش مصنوعی

تمرین هشتم

استاد:

مهدی سمیعی

نویسنده :

محمد هومان کشوری

شماره دانشجویی :

99105667

# تمرینات تئوری

## سوال 1.

(الف)

عملا این کار یعنی کل coefficients و وزن‌ها برابر  $\bullet$  قرار دهیم، عملکرد متفاوتی دارد اما غالبا توصیه نمی‌شود چرا که غالبا باعث مشکل در همگرایی می‌شود. معمولا برای رفع این مشکل مقادیر ابتدایی وزن‌ها را رندوم قرار می‌دهند تا کمترین همبستگی بین اعداد وجود داشته باشد.

(ب)

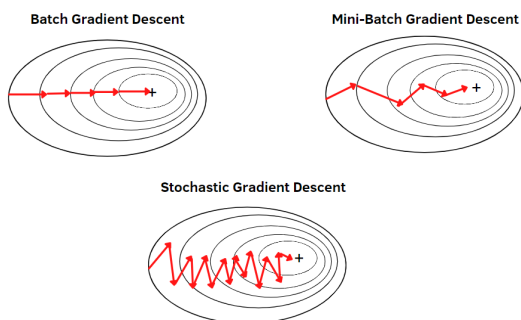
نیازی به شافل کردن نیست چرا که در full batch عملا روی کل داده‌ها داریم عملیات train کردن را انجام می‌دهیم؛ پس می‌توان گفت نهایتا در جواب نهایی چه با شافل و چه بدون شافل یک جواب خواهیم گرفت.

(ج)

بله، چرا که از بایاس شدن آموزش جلوگیری می‌کند و می‌توان گفت اگر این کار را نکنیم شاید یکی از آنها بهتر تشخیص داده شود

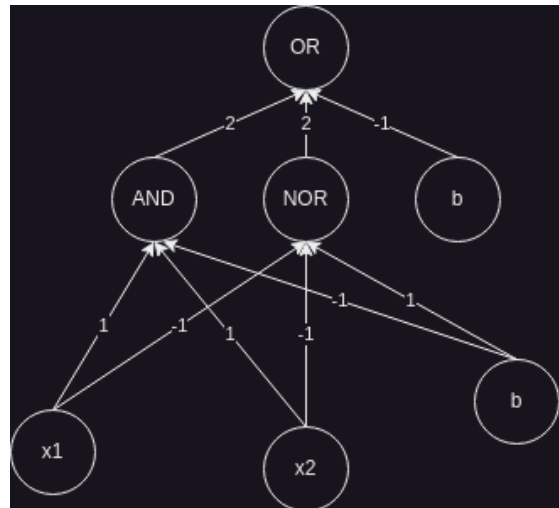
(د)

در مدل فول بچ چون روی تمام داده‌ها ترین را انجام می‌دهیم احتمالا خیلی ترین کردن نرم جلو رود (A). بر روی minibatch در فرایند ترین کردن تعداد کمتری داده داریم که یعنی کمی در پیدا کردن اپتیمال سولوشن تداخل داریم (B) و در stochastic بیشترین میزان تداخل را شاهد هستیم (C).



## سوال 2.

این سوال عملا برای دو بیت می‌خواهد XOR آن دو را پیدا کند.  
پس در لایه ورودی، دو input node قرار می‌دهیم که عملا هر کدام می‌تواند ۰ یا ۱ باشد  
که باید خروجی نهایتا XOR آنها را نمایش دهد.



حال این عبارت تساوی دو بیت  $x1$  ,  $x2$  را تشخیص می‌دهد (خروجی OR)  
حال می‌توان دو عدد ورودی را بیت به بیت همبستگی مقایسه کرد و نهایتا اگر خروجی  
کلی 0 بود جواب غلط است و در غیر این صورت درست است.

### سوال 3.

(الف)

در هر مرحله ۱ بایاس هم داریم که وزن آن برابر ۱ است و نباید آنرا فراموش کرد.  
 $(4096+1) * 2048 + 2 * (2048+1)$

(ب)

$$dJ/dW_{ij} = dJ/dz * dz/dW_{ij} = (z-y) * (h)$$

(ج)

$$dJ/dV_{ij} = dJ/dz * dz/dW_{ij} * dh/dg * dg/dV_{ij} = (z-y) * x$$

## سوال 4.

الف) درست، مطابق با توضیحات داده شده در صفحه ویکیپدیا، این روش زمان آموزش را کاهش می‌دهد.

ب) درست، با نرمال کردن ورودی‌ها تاثیر مقداردهی اولیه ضعیف وزن‌ها کاهش می‌یابد.

## سوال 5.

(الف)

۱. کاهش ابعاد به این صورت که می‌توانیم از آنان برای کاهش تعداد کانال‌ها استفاده کنیم.

۲. می‌توانیم از آنان برای حذف محدودیت خطی بودن در شبکه استفاده کنیم و به واسطه آنها از توابع غیر خطی در شبکه نیز استفاده کنیم.

۳. می‌توان از آنان برای ترکیب ویژگی‌ها در فیچر مپ استفاده کرد.

(ب)

به این صورت عمل می‌کنیم که برای پدینگ، باید پدینگ‌ها یکسان باشد که می‌توان از فرمول زیر استفاده کرد.

$$\text{Padding} = (\text{Filter size} - 1)/2$$

(ج)

تفاوت اصلی Max Pooling و Average Pooling در نحوه تجمیع اطلاعات در یک ناحیه مشخص از فیچر مپ ورودی است.

Max pooling بیشتر بر روی ویژگی‌های مهم ورودی تمرکز می‌کند اما average pooling تقریباً تمامی ویژگی‌ها را به صورت یکسان بررسی می‌کند.