



دانشکده‌ی مهندسی کامپیوتر

یادگیری ماشین

بهار ۱۴۰۰

تمرین سری چهارم

مدرس: دکتر محمدحسین رهبان

زمان تحویل: ۳ اردیبهشت

سوال ۱ MultiLayer Networks

- آ) با استفاده از *Perceptron* عملگرهای منطقی *NAND* و *NOR* را به گونه‌ای پیاده‌سازی کنید که هر کدام دو ورودی منفی یک (صفر منطقی) یا یک گرفته و خروجی منفی یک یا یک تولید کنند. (۴ نمره)
- ب) سه عدد حقیقی x و y و z را خوب می‌نامیم اگر داشته باشیم:

$$(x > y \text{ and } y < z) \text{ or } (x < y \text{ and } y > z)$$

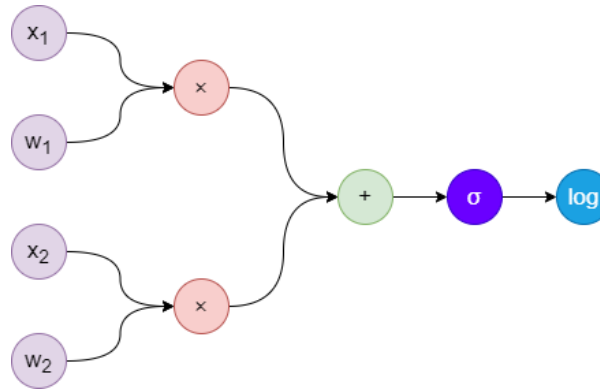
- با استفاده از *MultiLayer Perceptron (MLP)* شبکه‌ای طراحی کنید که سه عدد x ، y و z را به عنوان ورودی دریافت کرده و اگر خوب بودند خروجی یک و در غیر این صورت خروجی منفی یک تولید کند. (۸ نمره)

سوال ۲ Backpropagation & SGD

- آ) گراف محاسباتی شکل ۱ را در نظر بگیرید. به ازای مقادیر زیر و با استفاده از *Backpropagation* مشتق خروجی نهایی را نسبت به خروجی هر راس حساب کنید. (۶ نمره)

$$x_1 = 2, w_1 = 1, x_2 = -1, w_2 = 3$$

- ب) برای اعمال بهینه‌سازی می‌توان در هر *Epoch* به روش‌های مختلفی عمل کرد:
۱. همه داده‌ها را با هم در نظر گرفت و یک گام به ازای همه برداشت.
 ۲. به ازای هر داده (نقطه) یک گام جدا برداشت.
 ۳. داده‌های آموزش را به دسته‌های چندتایی تقسیم کرد و به ازای هر دسته یک گام برداشت.
- با بررسی این روش‌ها با توجه به هزینه محاسباتی، سرعت همگرایی، دقت و ... مزایا و معایب آن‌ها نسبت به یکدیگر را بیان کنید. (۶ نمره)



شکل ۱: گراف محاسباتی

سوال ۳ Regularization

در درس با $L2$ Regularization که تابع هدف را به شکل زیر تغییر می‌دهد آشنا شدید:

$$E = \left(y - \sum_i w_i x_i \right)^2 + \lambda \sum_i w_i^2$$

آ از عبارت بالا نسبت به وزن‌ها مشتق گرفته و با اعمال روش SGD توضیح دهید که چرا نام دیگر این روش $Weight Decay$ است. (۴ نمره)

ب) حال تابع هدف را به شکل زیر در نظر بگیرید:

$$E = \left(y - \sum_i w_i x_i \right)^2 + \lambda \sum_i |w_i|$$

با تکرار روند بخش آ برای تابع بالا توضیح دهید که چه تفاوتی در این روش با روش بخش آ وجود دارد؟ اگر وزنی کوچک باشد، چه اتفاقی برای آن می‌افتد؟ آیا این روش مزیتی هم دارد؟ (۸ نمره)

راهنمایی: دقت کنید که در خیلی از کاربردهای یادگیری ماشین، ابعاد ورودی بسیار بزرگ است ولی لزوماً به ازای هر ورودی، به همه ابعاد برای تولید خروجی نیاز نداریم.

سوال ۴ Validation

آ) توضیح دهید تفاوت داده‌های $test$ و $validation$ چیست؟ چرا از داده‌های $test$ برای $validation$ استفاده نمی‌کنیم؟ (۶ نمره)

ب) آیا در حالت کلی خطای مدل روی $validation$ set یک تخمینگر نااریب^۱ برای خطای $test$ set است؟ اگر از K -Fold Cross Validation استفاده کنیم چه؟ توضیح دهید. (۶ نمره)

^۱Unbiased

سوال ۵ (عملی) فایل *Notebook* ی که در اختیار شما قرار داده شده را کامل کنید. در این تمرین شما باید یک *MLP* روی داده‌های *MNIST* آموزش دهید تا بتواند برچسب عکس‌ها را پیش‌بینی کند. همچنین می‌توانید در انتهای *Notebook* بخشی اضافه کرده و طراحی شبکه را بهبود دهید. در صورت بهبود کارکرد شبکه و توضیح چرایی بهبود آن، نمره امتیازی متناسب برای شما در نظر گرفته می‌شود. (۴۰ نمره + ۱۵ نمره امتیازی)

پاینده باشید