



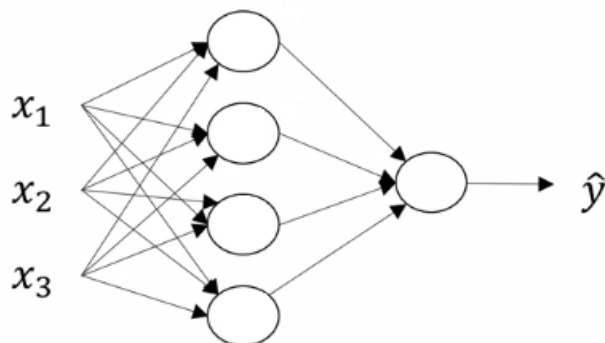
دانشگاه صنعتی اصفهان
دانشکده مهندسی برق و کامپیوتر

تمرین تئوری سری اول
مبانی هوش محاسباتی
زمستان ۱۴۰۳

استاد درس: دکتر سمانه حسینی
دستیار آموزشی: امین کافی
موعد تحویل تکلیف: ۲۱ اسفند ۱۴۰۳

۱ آموزش یک شبکه عصبی - ۴۰ نمره

این شبکه دارای سه ویژگی ورودی به نام‌های x_1, x_2, x_3 و یک خروجی \hat{y} است. معماری شبکه در شکل ۱ نشان داده شده است.



شکل ۱: یک شبکه‌ی عصبی ساده با یک لایه‌ی پنهان و سه ویژگی ورودی.

۱.۱ ساختار شبکه و تابع هزینه

۱. لایه‌ی ورودی: شامل سه نورون ورودی (x_1, x_2, x_3) .

۲. لایه‌ی پنهان: شامل چهار نورون است. هر نورون در این لایه مجموع وزندار ویژگی‌های ورودی را محاسبه کرده، مقدار بایاس را اضافه کرده و سپس تابع فعال‌سازی \tanh را اعمال می‌کند:

$$z_j^{(1)} = \sum_{i=1}^3 w_{ij}^{(1)} x_i + b_j^{(1)}, \quad a_j^{(1)} = \tanh(z_j^{(1)}) \quad \text{برای } j = 1, 2, 3, 4.$$

۳. لایه‌ی خروجی: شامل یک نورون است که مجموع وزندار خروجی‌های لایه‌ی پنهان را محاسبه کرده، مقدار بایاس را اضافه کرده و تابع فعال‌سازی Sigmoid را اعمال می‌کند:

$$z^{(2)} = \sum_{j=1}^4 w_j^{(2)} a_j^{(1)} + b^{(2)}, \quad \hat{y} = \sigma(z^{(2)}) = \frac{1}{1 + e^{-z^{(2)}}}.$$

۴. تابع هزینه: تابع هزینه‌ی J به عنوان میانگین binary cross-entropy در تمام داده‌های آموزشی تعریف می‌شود:

$$\mathcal{L} = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})). \quad (1)$$

نمادگذاری:

□ $w_{ij}^{(l)}$: وزن متصل‌کننده‌ی نورون i در لایه‌ی $l - 1$ به نورون j در لایه‌ی l .

□ $b_j^{(l)}$: مقدار بایاس برای نورون j در لایه‌ی l .

□ $z_j^{(l)}$: ورودی وزندار به نورون j در لایه‌ی l قبل از اعمال تابع فعال‌سازی.

□ $a_j^{(l)}$: خروجی نورون j در لایه‌ی l پس از اعمال تابع فعال‌سازی.

۲.۱ مقداردهی اولیه‌ی وزن‌ها و تولید داده

ابتدا، یک داده‌ی تصادفی با سه ویژگی و یک برچسب کلاس (0 یا 1) تولید می‌شود:

x	y
(0.23, 0.45, 0.67)	1

مقداردهی اولیه‌ی وزن‌ها و بایاس‌ها به صورت تصادفی انجام شده است:

$$\mathbf{w}^{(1)} = \begin{bmatrix} 0.45 & -0.12 & 0.78 & 0.72 \\ 0.05 & 0.35 & -0.22 & -0.85 \\ -0.55 & 0.11 & 0.67 & 0.45 \end{bmatrix}, \quad \mathbf{b}^{(1)} = \begin{bmatrix} 0.1 \\ -0.1 \\ 0.2 \\ -0.2 \end{bmatrix},$$

$$\mathbf{w}^{(2)} = \begin{bmatrix} 0.25 \\ -0.35 \\ 0.15 \\ -0.05 \end{bmatrix}, \quad \mathbf{b}^{(2)} = 0.3.$$

۳.۱ سوالات

۱. نشان دهید در صورت عدم استفاده از توابع فعال‌ساز، این شبکه تفاوتی با یک شبکه عصبی بدون لایه پنهان ندارد. (۱۰ نمره)

۲. مشتقات زیر را به کمک قاعده زنجیره‌ای محاسبه کنید. (۲۰ نمره)

(آ) مشتق تابع ضرر \mathcal{L} نسبت به خروجی شبکه \hat{y} :

$$\frac{\partial \mathcal{L}}{\partial \hat{y}}$$

(ب) مشتق تابع فعال‌سازی Sigmoid نسبت به ورودی آن:

$$\frac{\partial \hat{y}}{\partial z^{(2)}}$$

(ج) مشتق تابع فعال‌سازی \tanh نسبت به ورودی آن:

$$\frac{\partial a_j^{(1)}}{\partial z_j^{(1)}}$$

(د) مشتق تابع هزینه J نسبت به ورودی نورون خروجی $z^{(2)}$:

$$\frac{\partial J}{\partial z^{(2)}}$$

(ه) مشتق ضرر نسبت به وزن‌های لایه‌ی پنهان:

$$\frac{\partial \mathcal{L}}{\partial w_{ij}^{(1)}}$$

(و) مشتق ضرر نسبت به بایاس‌های لایه‌ی پنهان:

$$\frac{\partial \mathcal{L}}{\partial b_j^{(1)}}$$

(ز) مشتق ضرر نسبت به وزن‌های لایه‌ی خروجی:

$$\frac{\partial \mathcal{L}}{\partial w_j^{(2)}}$$

(ح) مشتق ضرر نسبت به بایاس در لایه‌ی خروجی:

$$\frac{\partial \mathcal{L}}{\partial b^{(2)}}$$

۳. با در نظر گرفتن مقداردهی اولیه شبکه و داده داده شده یک بار forward propagation و سپس backpropagation را انجام دهید و وزن‌های شبکه را با نرخ یادگیری ۰.۰۱ آموزش دهید. (۱۰ نمره)

۲ توزیع نامتوازن داده‌های آموزش - ۲۰ نمره

یکی از مشکلات رایج در در شبکه‌های عصبی زمانی رخ می‌دهد که داده‌ها توزیع مناسبی ندارند. به عنوان مثال، فرض کنید مجموعه داده‌ای داریم که شامل ۱۰۰۰۰۰۰ نمونه با برچسب ۱ و تنها ۱۰۰۰ نمونه با برچسب ۰ باشد. در این حالت، شبکه تمایل دارد که همیشه مقدار ۱ را خروجی دهد و در نتیجه تمامی داده‌های با برچسب ۰ را به اشتباه طبقه‌بندی می‌کند. برای حل این مشکل، یک راهکار ارائه دهید.

۳ اثر نرخ یادگیری، تابع هزینه و نقطه آغازین در نزول گرادیان - ۳۰ نمره

تابع هزینه‌ی زیر را در نظر بگیرید:

$$f(x) = x^4 + 4x^3 - 688x^2 + 2672x - 2640$$

می‌خواهیم به کمک روش Gradient Descent مقدار بهینه‌ی این تابع را پیدا کنیم. پنج گام از الگوریتم Gradient Descent را از مقدار اولیه‌ی $x_0 = 5.5$ طی کنید. یک‌بار با نرخ یادگیری $\alpha = 0.001$ و بار دیگر با $\alpha = 0.0026$.

۱. روند همگرایی روش Gradient Descent را با هر دو مقدار نرخ یادگیری بررسی کنید.

۲. تأثیر نرخ یادگیری، تابع هزینه و نقطه‌ی شروع را در یافتن مقدار بهینه تحلیل کنید.

۳. آیا انتخاب یک نرخ یادگیری مناسب همیشه باعث رسیدن به نقطه‌ی بهینه می‌شود یا این امر به تابع هزینه و مقدار اولیه نیز وابسته است؟ پاسخ خود را با استدلال توضیح دهید.

برای انجام محاسبات می‌توانید از ابزارهای آنلاین زیر کمک بگیرید:

Comprehensive Gradient Descent Visualizer □

UCLA ACM Gradient Descent Visualizer □

نیازی به ذکر محاسبات عددی نیست.

۴ مسئله دسته‌بندی چندکلاسه - ۱۰ نمره

در Logistic Regression کلاس‌های خروجی معمولاً دودویی (۰ یا ۱) در نظر گرفته می‌شوند. حال فرض کنید مسئله‌ی دسته‌بندی شامل سه کلاس ۰، ۱ یا ۲ باشد.

۱. آیا می‌توان از Logistic Regression برای این مسئله استفاده کرد؟ چگونه؟
۲. چگونه باید مدل را گسترش داد تا قادر به دسته‌بندی چندکلاسه باشد؟
۳. یک تابع هزینه مناسب برای این مدل پیشنهاد دهید و دلیل انتخاب خود را توضیح دهید.

۵ مقداردهی اولیه‌ی صفر برای وزن‌ها (امتیازی) - ۱۰ نمره

اگر تمام وزن‌های اولیه‌ی شبکه را برابر صفر قرار دهیم، چه مشکلی ممکن است پیش آید؟ آیا شبکه همچنان قادر به یادگیری خواهد بود؟ دلیل آن را توضیح دهید.