



مدرسین: دکتر فدایی و دکتر یعقوبزاده

طراحان: پرهام سازدار، مهدی جمالخواه

مهلت تحویل: سه‌شنبه ۱۸ دی ۱۴۰۳، ۲۳:۵۵

یادگیری ماشین

سوال اول: درخت تصمیم

یک رستوران تصمیم دارد بررسی کند که با توجه به عوامل موثر، در صورتی که تمام میزها پر باشد، آیا افرادی که به رستوران مراجعه می‌کنند برای خالی شدن میز صبر می‌کنند یا خیر.

Example	Input Attributes										Goal
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	WillWait
x ₁	Yes	No	No	Yes	Some	\$\$\$	No	Yes	French	0-10	y ₁ = Yes
x ₂	Yes	No	No	Yes	Full	\$	No	No	Thai	30-60	y ₂ = No
x ₃	No	Yes	No	No	Some	\$	No	No	Burger	0-10	y ₃ = Yes
x ₄	Yes	No	Yes	Yes	Full	\$	Yes	No	Thai	10-30	y ₄ = Yes
x ₅	Yes	No	Yes	No	Full	\$\$\$	No	Yes	French	>60	y ₅ = No
x ₆	No	Yes	No	Yes	Some	\$\$	Yes	Yes	Italian	0-10	y ₆ = Yes
x ₇	No	Yes	No	No	None	\$	Yes	No	Burger	0-10	y ₇ = No
x ₈	No	No	No	Yes	Some	\$\$	Yes	Yes	Thai	0-10	y ₈ = Yes
x ₉	No	Yes	Yes	No	Full	\$	Yes	No	Burger	>60	y ₉ = No
x ₁₀	Yes	Yes	Yes	Yes	Full	\$\$\$	No	Yes	Italian	10-30	y ₁₀ = No
x ₁₁	No	No	No	No	None	\$	No	No	Thai	0-10	y ₁₁ = No
x ₁₂	Yes	Yes	Yes	Yes	Full	\$	No	No	Burger	30-60	y ₁₂ = Yes

جدول ۱-۱: دادگان جمع‌آوری شده از مراجعان

توضیح ویژگی‌های مختلف به صورت زیر است:

Alternate: whether there is a suitable alternative restaurant nearby.
Bar: whether the restaurant has a comfortable bar area to wait in.
Fri/Sat: true on Fridays and Saturdays.
Hungry: whether we are hungry.
Patrons: how many people are in the restaurant (values are None, Some, and Full).
Price: the restaurant's price range (\$, \$\$, \$\$\$).
Raining: whether it is raining outside.
Reservation: whether we made a reservation.
Type: the kind of restaurant (French, Italian, Thai or Burger).
WaitEstimate: the wait estimated by the host (0-10 minutes, 10-30, 30-60, >60).

جدول ۱-۲: توضیح ویژگی‌ها

درخت تصمیم مناسب برای پیش‌بینی اینکه مشتری‌ها صبر می‌کنند یا خیر را به دست بیاورید. تمامی مراحل انتخاب ویژگی برای درخت تصمیم را بنویسید.

شبکه عصبی

سوال اول: Backpropagation Basics

فرض کنید شبکه عصبی با دو لایه مانند زیر داریم:

$$z_1 = W_1 x^{(i)} + b_1$$

$$a_1 = \text{ReLU}(z_1)$$

$$z_2 = W_2 a_1 + b_2$$

$$\hat{y}^{(i)} = \sigma(z_2)$$

$$L^{(i)} = y^{(i)} * \log(\hat{y}^{(i)}) + (1 - y^{(i)}) * \log(1 - \hat{y}^{(i)})$$

$$J = \frac{1}{m} \sum_{i=1}^m L^{(i)}$$

توجه کنید که $x^{(i)}$ نشان دهنده یک نمونه ورودی با ابعاد $1 \times D_{in}$ است. همچنین $y^{(i)}$ برچسب یک نمونه است و به صورت اسکالر می باشد. دیتاست شامل m نمونه است. همچنین z_1 ابعاد $1 \times D_1$ دارد.

- ابعاد W_1, b_1, W_2, b_2 را بنویسید.
- نتیجه $\partial J / \partial y^{(i)}$ را بدست آورید و آن را با δ_1 نشان دهید.
- نتیجه $\partial y^{(i)} / \partial z_2$ را بدست آورید و آن را با δ_2 نشان دهید.
- نتیجه $\partial z_2 / \partial a_1$ را بدست آورید و آن را با δ_3 نشان دهید.
- نتیجه $\partial a_1 / \partial z_1$ را بدست آورید و آن را با δ_4 نشان دهید.
- نتیجه $\partial z_1 / \partial W_1$ را بدست آورید و آن را با δ_5 نشان دهید.
- نتیجه $\partial J / \partial W_1$ را با استفاده از نتایج قبلی بدست آورید.

یادآوری: در محاسبات ماتریسی، مشتق به صورت زیر تعریف می شود:

- مشتق بردار نسبت به بردار

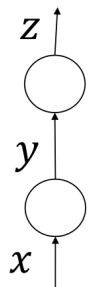
$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix} = f \left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \right) \Rightarrow \frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_d}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_d}{\partial x_2} \\ \cdots & \cdots & \ddots & \cdots \\ \frac{\partial y_1}{\partial x_m} & \frac{\partial y_2}{\partial x_m} & \cdots & \frac{\partial y_d}{\partial x_m} \end{bmatrix}$$

- مشتق بردار نسبت به ماتریس

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix} = f \left(\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \ddots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \right) \Rightarrow \frac{\partial y_1}{\partial X} = \begin{bmatrix} \frac{\partial y_1}{\partial x_{11}} & \frac{\partial y_1}{\partial x_{12}} & \cdots & \frac{\partial y_1}{\partial x_{1n}} \\ \frac{\partial y_1}{\partial x_{21}} & \frac{\partial y_1}{\partial x_{22}} & \cdots & \frac{\partial y_1}{\partial x_{2n}} \\ \cdots & \cdots & \ddots & \cdots \\ \frac{\partial y_1}{\partial x_{m1}} & \frac{\partial y_1}{\partial x_{m2}} & \cdots & \frac{\partial y_1}{\partial x_{mn}} \end{bmatrix}$$

همچنین در مشتق بردارها نیز قانون زنجیره‌ای (chain rule) برقرار است: (دقت کنید که ابتدا مشتق بیرونی نوشته می‌شود و بعد مشتق درونی. اگر برعکس نوشته شود، اشتباه است؛ زیرا در ضرب ماتریس‌ها قانون جابه‌جایی برقرار نیست)

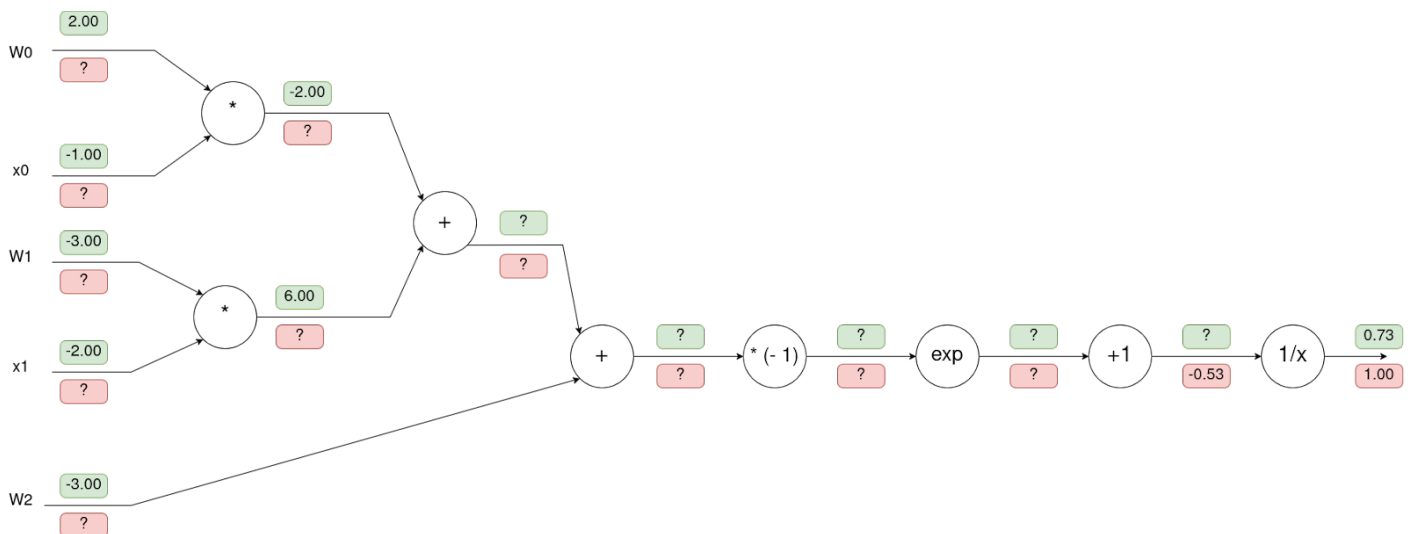
- $z = f(g(x))$
- $y = g(x)$



$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} \frac{\partial y}{\partial x}$$

سوال دوم: Forward & Backward

عبارت ریاضی $f(w, x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2x_2)}}$ را در نظر بگیرید که در شکل زیر گراف محاسباتی آن نشان داده شده است. بالای هر یال مقدار Forward و پایین آن مقدار Backward که همان مشتق نسبت به متغیر ورودی یال است نشان داده شده است. بعضی از این مقادیر با علامت (?) جایگذاری شده است. در پاسخ‌تان همین گراف را رسم کنید و مقادیر مجهول را با مقدار درست آن‌ها جایگذاری کنید. روابط ریاضی‌ای که در محاسبه این مقادیر استفاده می‌کنید نیز بنویسید.



سوال سوم: CNN

به سوالات زیر در حد یک یا دو خط پاسخ کوتاه دهید.

الف) فرض کنید برداری به طول N دارید و قصد دارید یک لایه کانولوشن یک بعدی روی آن اعمال کنید. حاصل اعمال یک لایه کانولوشن را از طریق رابطه‌ی:

$$z = w * x \rightarrow Z_i = \sum_{j=0}^{K-1} w_j x_{i+j}$$

بدست می‌آوریم که K اندازه فیلتر را نشان می‌دهد. فرض کنید به سمت راست بردار x ، $(K - 1)$ تا صفر به عنوان Padding اضافه می‌کنیم. اگر مقدار $\frac{\partial Loss}{\partial Z_i}$ را برای تمامی مقادیر i بدانیم، رابطه محاسبه $\frac{\partial Loss}{\partial w_j}$ را به طور دقیق بر حسب آن پیدا کنید. نشان دهید که این رابطه برای پیدا کردن $\frac{\partial Loss}{\partial w_j}$ عملاً معادل اعمال یک فیلتر کانولوشن است.

ب) مرتبه‌ی تعداد عملیات محاسباتی لازم برای یک لایه کانولوشن با اندازه‌ی کرنل $k \times k$ و گام s ، به ازای یک ورودی به عرض w و ارتفاع h که دارای c کانال است بنویسید.

ج) یک بلوک CNN به صورت زیر را در نظر بگیرید:

3x3 Conv (stride 2) - 2x2 Pool (stride 2) - 3x3 Conv (stride 2) - 2x2 Pool (stride 2)

حال **receptive field** یک پیکسل خروجی این بلوک را بدست آورید.