



## آزمون پایان ترم

- زمان در نظر گرفته شده برای آزمون ۱۵۰ دقیقه است.
- لطفا پاسخ‌های خود را خوانا و خوش خط بنویسید.
- پاسخ هر سوال باید در یک برگه جداگانه نوشته شود. بالای هر برگه ی پاسخ نامه، نام و شماره دانشجویی خود را به صورت واضح بنویسید.

## سوالات (۵+۱۰۰ نمره)

۱. (۲۰ نمره) به سوالات زیر به طور مختصر پاسخ دهید:

- (آ) لایه pooling به دلیل نداشتن وزنی برای یادگیری تأثیری در backpropagation ندارد. این عبارت درست است یا غلط؟ با ذکر دلیل مشخص کنید.
- (ب) یک شبکه عصبی fully connected را در نظر بگیرید که تابع فعالسازی تمام لایه‌ها تابع tanh می‌باشد. برای مقداردهی اولیه وزن‌ها، همه وزن‌های شبکه را مقادیری بزرگ انتخاب می‌کنیم. این روش ایده خوبی برای کارکرد این شبکه عصبی نیست. این عبارت درست است یا غلط؟ با ذکر دلیل مشخص کنید. راهنمایی:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

(ج) اشکال تخمین زدن احتمال  $P(A|B)$  به صورت  $\frac{\text{تعداد رخداد A و B با یکدیگر}}{\text{تعداد رخداد B}}$  چیست؟ برای بهبود آن چه روشی را پیشنهاد می‌کنید؟

(د) توضیح دهید هرکدام از روش‌های زیر چه تأثیری در بیش‌برازش<sup>۱</sup> کردن مدل‌ها دارند.

- کم کردن تعداد برگ‌ها در درخت تصمیم
- محدود کردن حداکثر طول درخت تصمیم

درباره تأثیر آن‌ها در دقت مدل در داده‌های آموزش چه می‌توان گفت؟

(ه) عبارت زیر درست است یا غلط؟ با ذکر دلیل مشخص کنید. یک شبکه بیزی داریم که در آن  $X$  به شرط  $Z$  از  $Y$  مستقل است. ممکن است فرض استقلال این دو متغیر با شرطی کردن شواهد اضافه برای متغیرهای دیگر در شبکه، برقرار نباشد.

۲. (۱۵ نمره) می‌خواهیم یک شبکه عصبی طراحی کنیم که عبارت زیر را پیاده سازی کند:

$$(x \vee \neg y) \oplus (\neg m \vee \neg n)$$

(علامت  $\neg$  معادل not منطقی، علامت  $\oplus$  معادل xor منطقی و علامت  $\vee$  معادل or منطقی است.) در طراحی شبکه‌های عصبی تابع فعال‌سازی را به صورت زیر در نظر بگیرید:

$$f(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}$$

برای این منظور ابتدا برای عبارت‌های زیر شبکه عصبی طراحی کنید:

<sup>۱</sup>Overfit

- $(x \vee \neg y)$
- $(\neg m \vee \neg n)$
- $A \oplus B$

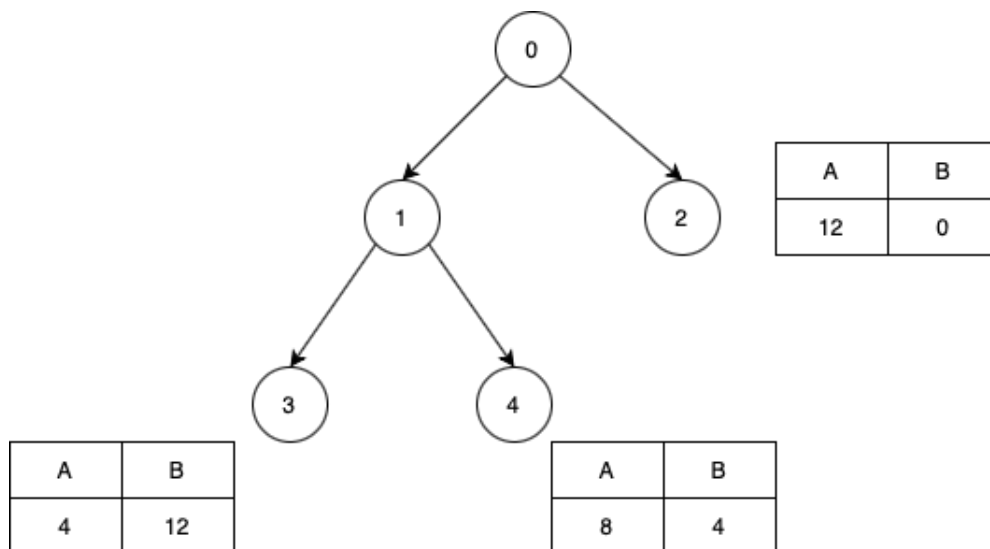
سپس با ترکیب شبکه‌های به دست آمده برای عبارت گفته‌شده شبکه عصبی طراحی کنید. (توجه کنید که وزن‌ها و بایاس‌ها را به صورت دقیق مشخص کنید.)

۳. (۲۰ نمره) مسئله‌ی رگرسیون خطی  $\hat{y} = w^T x$  را برای مجموعه داده‌ی  $D = (x_i, y_i)_{i=1}^n$  و با تابع زیان  $J(w) = \sum_{i=1}^n (y^{(i)} - \hat{y}^{(i)})^2$  در نظر بگیرید.

(آ) ابتدا  $\operatorname{argmin}_w J(w)$  را به دست آورده و ساده کنید.

(ب) فرمول به دست آمده در قسمت قبل ممکن است چه مشکلاتی را به همراه داشته باشد؟ در این قسمت  $\operatorname{argmin}_w J(w) + \lambda \|w\|^2$  را بدست آورید و ساده کنید. سپس توضیح دهید این عبارت چگونه مشکل قسمت قبل را حل می‌کند.

۴. (۱۰ نمره) درخت تصمیم زیر را در نظر بگیرید:

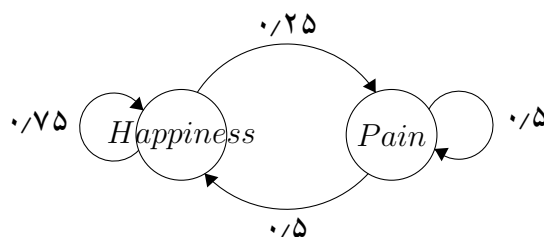


الف) آنتروپی و میزان اطلاعات به دست آمده بر اساس گره شماره صفر را به دست آورید.

ب) معیار دقت را برای classification انجام شده بر روی این درخت تصمیم به دست آورید.

پ) پیشنهاد شما برای افزایش این معیار چیست؟ راه حل پیشنهادی شما ممکن است چه مشکلاتی را برای این درخت تصمیم ایجاد کند و راهکار جایگزین شما برای این مورد چیست؟

۵. (۲۰ نمره) markov chain زیر دو استیت گلی (شاد و غمگین) را نشان می‌دهد. در ابتدا در حالت شادی هستیم.



فرض کنید یک سری از استیت‌های گلی به صورت دنباله  $x_1, x_2, x_3, \dots, x_n$  از این markov chain نمونه‌برداری شده است. ما می‌توانیم گروه‌های این نمونه را تشکیل بدهیم. برای مثال:

$H, H, H, P, P, H, H, P, P, P, P, P$  شامل چهار گروه با اندازه‌های ۳، ۲، ۲، ۵ است.

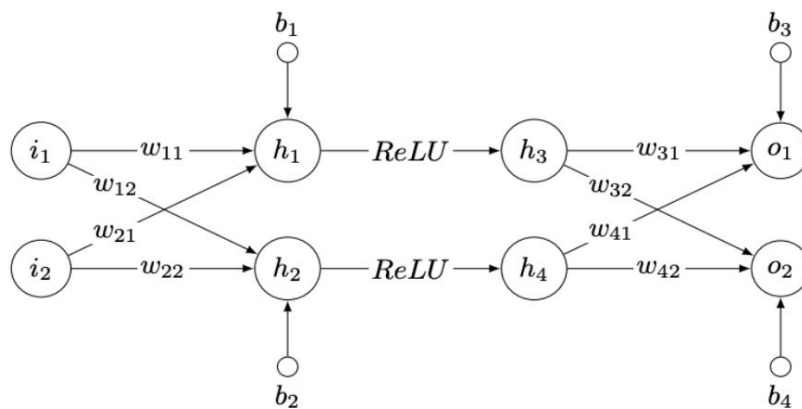
$G_n$  به صورت زیر تعریف می‌شود:

$$G_n = \frac{n}{\#groups}$$

برای مثال در سوال ما  $G_n = \frac{12}{4} = 3$  می‌شود.

گلی ادعا کرده است با پیدا کردن مقداری که  $G_n$  به آن همگرا می‌شود همیشه در استیت خوشحالی خواهد ماند. مقداری که  $G_n$  به آن همگرا می‌شود را پیدا کنید.

۶. (۲۰ نمره) شبکه عصبی شکل ۱ را با تابع فعال‌سازی ReLU در نظر بگیرید.  $(i_1, i_2)$  ورودی هستند، دو لایه مخفی داریم و خروجی‌ها در انتها  $(o_1, o_2)$  هستند. برجسب داده‌ها با  $(t_1, t_2)$ ، وزن‌ها با  $w$  و بایاس با  $b$  نشان داده شده است.



شکل ۱: شبکه عصبی

مقادیر متغیرها را هم می‌توانید در جدول شکل ۲ مشاهده کنید.

Variable	$i_1$	$i_2$	$w_{11}$	$w_{12}$	$w_{21}$	$w_{22}$	$w_{31}$	$w_{32}$	$w_{41}$	$w_{42}$	$b_1$	$b_2$	$b_3$	$b_4$	$t_1$	$t_2$
Value	2.0	-1.0	1.0	-0.5	0.5	-1.0	0.5	-1.0	-0.5	1.0	0.5	-0.5	-1.0	0.5	1.0	0.5

شکل ۲: جدول مقادیر متغیرها

(آ) خروجی  $(o_1, o_2)$  را با توجه به مقادیر داده شده به دست بیاورید. تمامی محاسبات را بنویسید.

(ب) خطای  $MSE$  را حساب کنید.

(ج) فرض کنید تابع هزینه همان قسمت ب باشد. مقدار وزن  $w_{21}$  را با کمک gradient descent با نرخ یادگیری ۰/۱ آپدیت کنید. (تمامی محاسبات را بنویسید)

راهنمایی:

$$ReLU(x) = \max(0, x)$$