



هوش مصنوعی

بهار ۱۴۰۴

استاد: احسان تن قطاری

دانشگاه صنعتی شریف

دانشکده مهندسی کامپیوتر

طراحان: سپهر غیاث، امیرکوشان فتاح، شایان بقایی نژاد، آرش ضیائی، رادین چراغی، علیرضا ملک حسینی

مهلت ارسال: ۳۰ اردیبهشت

یادگیری ماشین

تمرین چهارم

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- در طول ترم امکان ارسال با تاخیر پاسخ همه‌ی تمارین سقف ۴ روز و در مجموع ۱۰ روز، وجود دارد. پس از گذشت این مدت، پاسخ‌های ارسال شده پذیرفته نخواهند بود. همچنین، به ازای هر ساعت تأخیر غیر مجاز نیم درصد از نمره‌ی تمرین کم خواهد شد.
- همکاری و هم‌فکری شما در انجام تمرین مانعی ندارد اما پاسخ ارسالی هر کس حتما باید توسط خود او نوشته شده باشد.
- در صورت هم‌فکری و یا استفاده از هر منابع خارج درسی، نام هم‌فکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.
- لطفا تصویری واضح از پاسخ سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.

سوالات نظری (۱۵+۱۰۰ نمره)

۱. (۵ نمره) درستی یا نادرستی عبارت‌های زیر را با ذکر دلیل مشخص کنید.
 - (آ) افزایش پیچیدگی یک مدل رگرسیون، همواره منجر به کاهش خطای مدل در داده‌های آموزش و افزایش خطای مدل در برآورد داده تست می‌شود.
 - (ب) اگر bias مدل ما زیاد است، اضافه کردن تعداد داده‌های آموزش کمک زیادی به کم کردن bias نمی‌کند.
 - (ج) از آنجا که طبقه بندی حالت خاصی از رگرسیون است، رگرسیون لجیستیک نیز حالت خاصی از رگرسیون خطی می‌باشد.
 - (د) در صورتی که یک نمونه از داده‌های آموزشی توسط الگوریتم پرسپترون اشتباه کلاس بندی شود، پس از بروزرسانی وزن‌های مدل با استفاده از همان نمونه، الگوریتم آنرا به درستی کلاس بندی می‌کند.
۲. (۱۲ نمره) شما تصمیم گرفته‌اید معلم شوید. تنها مشکلی که وجود دارد این است که نمی‌خواهید زمان زیادی را صرف تصحیح مقاله‌ها کنید، بنابراین تصمیم می‌گیرید همه آن‌ها را با یک طبقه‌بند خطی نمره‌دهی کنید. طبقه‌بند شما تعداد کلمات ۷ حرفی (f_7) و ۸ حرفی (f_8) در یک مقاله را در نظر می‌گیرد و سپس براساس این دو عدد به مقاله نمره A یا F می‌دهد. شما چهار مقاله نمره‌دار برای یادگیری در اختیار دارید:

تعداد کلمات ۷ حرفی (f_7)	تعداد کلمات ۸ حرفی (f_8)	نمره (A (+), F (-))
۳	۲	A
۱	۱	F
۴	۳	A
۰	۲	F

- (آ) ابتدا با ذکر دلیل مشخص کنید که داده‌های آموزشی با ویژگی‌های داده‌شده به صورت خطی جداپذیر هستند یا نه.

(ب) شما تصمیم می‌گیرید الگوریتم پرسپترون را اجرا کنید و با خوش‌بینی نسبت به توانایی‌های نگارش مقاله‌ی دانش‌آموزان، بردار وزن خود را به صورت $(1, 0, 0)$ مقداردهی اولیه می‌کنید. اگر امتیاز خروجی طبقه‌بند شما بیشتر از ۰ باشد، نمره‌ی A می‌دهد؛ اگر ۰ یا کمتر باشد، نمره‌ی F می‌دهد. بردار وزن حاصل را پس از مشاهده‌ی اولین و دومین مثال آموزشی با استفاده از الگوریتم پرسپترون بنویسید.

(ج) (۴ نمره) برای هر یک از قوانین تصمیم‌گیری زیر، مشخص کنید آیا بردار وزنی وجود دارد که آن قانون را نمایش دهد. اگر پاسخ «بله» است، بردار وزنی مربوطه را بنویسید.

- مقاله نمره A می‌گیرد اگر و تنها اگر $(f_v + f_h \geq 7)$ برقرار باشد.
- مقاله نمره A می‌گیرد اگر و تنها اگر $(f_v \geq 5 \text{ و } f_h \geq 4)$ برقرار باشد.
- مقاله نمره A می‌گیرد اگر و تنها اگر $(f_v \geq 5 \text{ یا } f_h \geq 4)$ برقرار باشد.
- مقاله نمره A می‌گیرد اگر و تنها اگر بین ۴ تا ۶ (شامل هر دو) کلمه ۷ حرفی و بین ۳ تا ۵ کلمه ۸ حرفی داشته باشد.

۳. (۱۵ نمره) فرض کنید یک مجموعه داده با سه ویژگی ورودی دودویی به صورت زیر A, B, C و یک ویژگی خروجی دودویی Y در اختیار داریم.

ویژگی‌های ورودی مقادیر $\{0, 1\}$ می‌گیرند، در حالی که Y مقادیر $\{\text{نادرست}, \text{درست}\}$ را می‌گیرد.

Y	C	B	A
درست	۱	۱	۰
درست	۰	۱	۱
نادرست	۱	۰	۱
نادرست	۱	۱	۱
درست	۱	۱	۰
درست	۰	۰	۰
نادرست	۱	۱	۰
نادرست	۱	۰	۱
درست	۰	۱	۰
درست	۱	۱	۱

با توجه به اطلاعات ذکر شده، به موارد زیر پاسخ دهید.

- (آ) Naive Bayes رکورد $(A = 1, B = 1, C = 0)$ را چگونه طبقه‌بندی می‌کند؟
- (ب) Naive Bayes رکورد $(A = 0, B = 0, C = 1)$ را چگونه طبقه‌بندی می‌کند؟
- (ج) Naive Bayes رکورد $(A = 0, B = 0, C = 0)$ را با استفاده از Laplace smoothing ($k=1$) چگونه طبقه‌بندی می‌کند؟
- (د) آیا افزودن تنها یک رکورد می‌تواند باعث تغییر طبقه‌بندی رکورد $(A = 1, B = 0, C = 1)$ شود؟

۴. (۱۶ نمره) به پرسش‌های زیر با استفاده از مجموعه داده سوال قبل و جدول آنتروپی زیر پاسخ دهید.

Specific Conditional Entropies		
$H(Y A = 1, C = 0) = 0.00$	$H(Y A = 0, B = 0) = 0.00$	$H(Y A = 0) = 0.72$
$H(Y A = 1, C = 1) = 0.81$	$H(Y A = 0, B = 1) = 0.81$	$H(Y A = 1) = 0.97$
$H(Y B = 0, C = 0) = 0.00$	$H(Y A = 1, B = 0) = 0.00$	$H(Y B = 0) = 0.92$
$H(Y B = 0, C = 1) = 0.00$	$H(Y A = 1, B = 1) = 0.92$	$H(Y B = 1) = 0.86$
$H(Y B = 1, C = 0) = 0.00$	$H(Y A = 0, C = 0) = 0.00$	$H(Y C = 0) = 0.00$
$H(Y B = 1, C = 1) = 0.97$	$H(Y A = 0, C = 1) = 0.92$	$H(Y C = 1) = 0.99$

(آ) کدام یک از A, B, C بیشترین information gain را دارد.

(ب) درخت تصمیم را بدون استفاده از عملیات هرس برای این سوال رسم کنید. در هنگام ساخت درخت، اگر در مرحله‌ای به دو ویژگی یکسان برای تقسیم برخوردید، ویژگی را براساس ترتیب حروف الفبا انتخاب کنید. همچنین برای برچسب‌گذاری برگ‌ها، از True در صورت تساوی استفاده کنید.

(ج) خروجی درخت تصمیم شما برای ورودی‌های زیر چیست؟

(1) $(A = 0, B = 0, C = 1)$

(2) $(A = 1, B = 0, C = 0)$

(د) اگر تمامی رئوس به جز ریشه را از درخت تصمیم هرس کنید و فرض کنید در صورت تساوی True را انتخاب کنید. خروجی برای ورودی‌های زیر چه خواهد بود؟

(1) $(A = 0, B = 0, C = 1)$

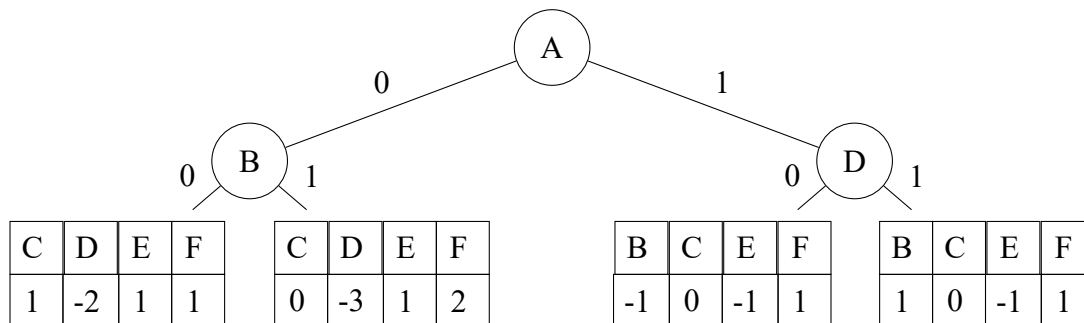
(2) $(A = 1, B = 0, C = 0)$

۵. (۱۲ نمره)

نیکا که به دنبال دسته‌بندی کننده‌ی مناسب خودش می‌گشت، الگوریتم جدیدی به نام «درخت پرسپترون» به ذهنش رسید که ویژگی‌هایی را از هر دو مدل درخت تصمیم و الگوریتم پرسپترون را باهم آمیخته می‌کند. درخت پرسپترون مشابه درخت تصمیم است، با این تفاوت که در هر برگ درخت به جای رای اکثریت^۱ از یک مدل پرسپترون استفاده می‌شود.

برای ساخت درخت پرسپترون، نخستین گام دنبال کردن روند عادی یک الگوریتم درخت تصمیم و جدا کردن بر اساس ویژگی‌هاست تا زمانی که به حداکثر عمق مجاز درخت برسیم. پس از رسیدن به این عمق، در هر برگ یک مدل پرسپترون بر پایه‌ی ویژگی‌های استفاده شده در آن شاخه یاد گرفته می‌شود. در نتیجه با داشتن یک نمونه‌ی جدید، ابتدا نمونه را وارد درخت تصمیم خود می‌کنیم و با پیش‌روی به یکی از برگ‌ها می‌رسیم. سپس مدل پرسپترون آن برگ را بر روی ویژگی‌های بررسی نشده انجام می‌دهیم و دسته‌بندی می‌کنیم.

فرض کنید که یک dataset با ۶ ویژگی دودویی^۲ $\{A, B, C, D, E, F\}$ و دو دسته‌ی خروجی $\{1, -1\}$ دارید. یک درخت پرسپترون با عمق ۲ برای این dataset در شکل زیر داده شده است. وزن‌های پرسپترون در هر برگ داده شده است. فرض کنید مقدار bias برای هر برگ $b = 1$ است.



(آ) خروجی درخت پرسپترون برای نمونه‌ی $x = \{1, 1, 0, 1, 0, 1\}$ چیست؟

(ب) درستی یا نادرستی عبارت‌های زیر را درباره‌ی درخت پرسپترون مشخص کنید.

۱. مرز تصمیم‌گیری درخت پرسپترون همواره خطی است.

۲. برای مقادیر کوچک حداکثر عمق (۲-۳)، احتمال underfit شدن درخت پرسپترون بیشتر از درخت تصمیم است.

^۱ Majority Voting
^۲ binary

ج) فرض کنید \mathcal{D} که یک dataset است، به ما داده شده است. از دو ساختار درخت تصمیم متفاوت برای یادگیری این dataset استفاده می‌کنیم. یک بار از Information Gain و بار دیگر از Training Error Rate برای جدا کردن نمونه‌ها استفاده می‌کنیم (در حالت دوم در هر گام بر اساس ویژگی‌ای تقسیم‌بندی می‌کنیم که کم‌ترین Training Error را می‌دهد). با فرض این که هر دو درخت تا رسیدن به Training Error صفر آموزش داده می‌شوند، کدام موارد زیر درست هستند؟

- هر دو درخت در ریشه بر پایه‌ی ویژگی یکسانی تقسیم‌بندی می‌کنند.
- هر دو درخت عمق یکسان خواهند داشت.
- هر دو درخت برای هر Data Point در \mathcal{D} خروجی یکسان خواهند داشت.
- هر دو درخت برای هر ورودی، خروجی یکسان خواهند داشت.
- هر دو درخت از تمام ویژگی‌های درون Dataset استفاده خواهند کرد.

۶. (۲۰ نمره) یکی از روش‌های منظم‌سازی در مسائل رگرسیون خطی، روش Lasso است. در این روش، نرم L_1 وزن‌های مدل در تابع خطا وارد می‌شود. این کار باعث می‌شود که پاسخ نهایی مسئله به صورت پراکنده‌تری (sparse) باشد. در این مسئله، خواهیم دید که چگونه جمله‌ی نرم L_1 منجر به افزایش پراکندگی می‌شود. هدف پراکنده کردن جواب نهایی (تنگ کردن آن) یا همون sparse کردن آن بهینه کردن فضای ذخیره سازی داده ها، ساده کردن محاسبات بعدی و جلوگیری از بیش برازش (overfitting) هست. $\mathbf{X} \in \mathbb{R}^{n \times d}$ ماتریسی هست که هر سطر آن یک مشاهده از d ویژگی می باشد و در کل n مشاهده داریم. $\mathbf{y} \in \mathbb{R}^n$ بردار برچسب ما می باشد. حال فرض کنید که $\mathbf{w} \in \mathbb{R}^d$ بردار وزن مدل رگرسیون ما می باشد و w^* وزن های بهینه می باشند. همچنین فرض کنید مشاهدات ما دارای خاصیت $\mathbf{X}^T \mathbf{X} = \mathbf{I}$ می باشند. (سفید شده اند) در رگرسیون لاسو بردار وزن های بهینه به صورت زیر به دست میاید :

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} J_{\lambda}(\mathbf{w})$$

$$J_{\lambda} = \frac{1}{2} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_1$$

(آ) ابتدا نشان می‌دهیم که سفیدسازی داده‌ها باعث مستقل شدن ویژگی‌ها می‌شود، به طوری که w_i^* تنها از ویژگی i ام نتیجه‌گیری می‌شود. برای اثبات این موضوع، ابتدا نشان دهید که J_{λ} را می‌توان به صورت زیر نوشت:

$$J_{\lambda}(\mathbf{w}) = g(\mathbf{y}) + \sum_{i=1}^d f(\mathbf{X}_{:,i}, \mathbf{y}, w_i, \lambda)$$

که در آن $\mathbf{X}_{:,i}$ ستون i ام ماتریس \mathbf{X} است.

(ب) اگر $w_i \geq 0$ باشد، مقدار w_i را بیابید.

(ج) اگر $w_i < 0$ باشد، مقدار w_i را بیابید.

(د) با توجه به بخش‌های قبلی، تحت چه شرایطی w_i برابر صفر خواهد شد؟ این شرایط چگونه قابل اعمال هستند؟

(ه) همان‌طور که می‌دانیم، در رگرسیون ریج، جمله‌ی منظم‌سازی در تابع هزینه به صورت $\frac{1}{2} \lambda \|\mathbf{w}\|_2^2$ ظاهر می‌شود. در این حالت، w_i تحت چه شرایطی برابر صفر می‌شود؟ تفاوت این حالت با حالت قبلی چیست؟

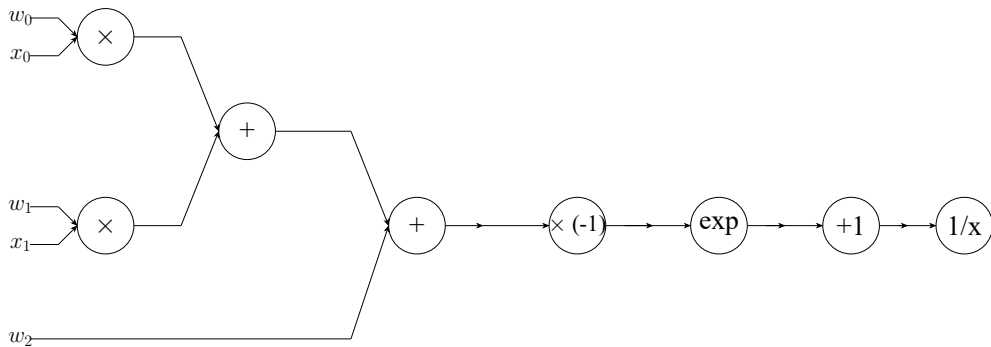
۷. (۱۵ نمره) تابع زیر را در نظر بگیرید :

$$f(w, x) = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2 x_2)}}$$

شبکه متناظر با این تابع در زیر آمده است. با فرض ورودی های

$$w_0 = 2, x_0 = -1, w_1 = -3, x_1 = -2, w_2 = -3$$

ابتدا خروجی نهایی را محاسبه کنید. سپس با استفاده از Backpropagation مشتق خروجی نهایی را نسبت به خروجی هر راس حساب کنید.



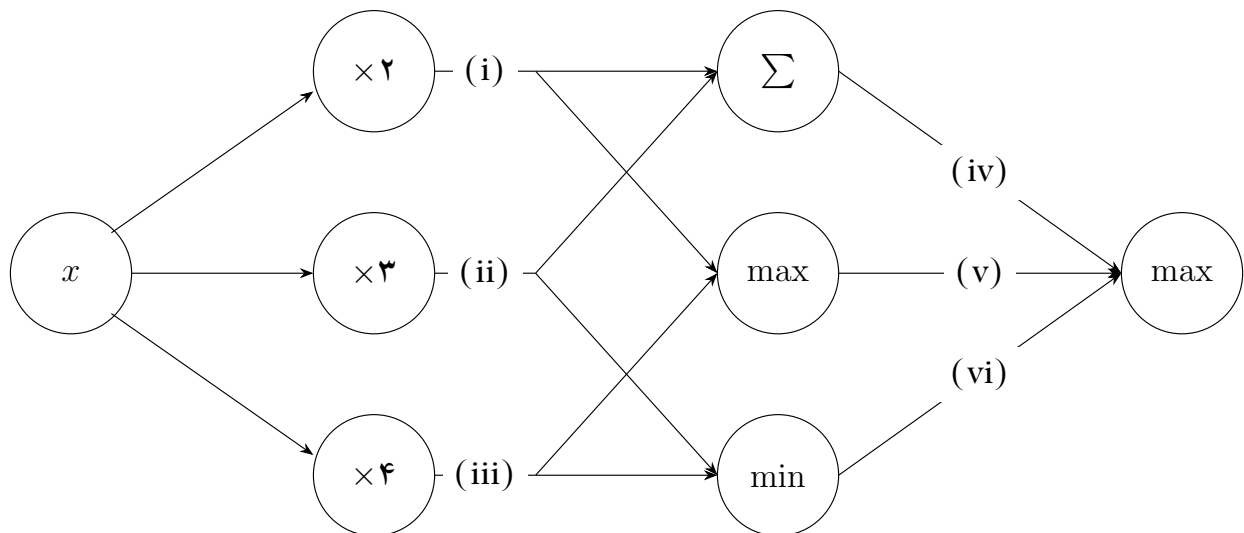
۸. (۵+۱۵ نمره)

(آ) عملیات Forward Propagation را برای شبکه‌ی عصبی زیر با ورودی $x = 1$ انجام دهید.

۱) مقادیر مربوط به (i), (ii), (iii) را به دست آورید.

۲) مقادیر مربوط به (iv), (v), (vi) را به دست آورید.

۳) مقدار خروجی مدل را به دست آورید.



(ب) (امتیازی) شکل زیر یک شبکه‌ی عصبی با وزن‌های a, b, c, d, e, f و d را نشان می‌دهد. ورودی این شبکه x_1 و x_2 هستند. لایه‌ی مخفی نخست به شکل زیر محاسبه می‌شود.

$$r_1 = \max(c \cdot x_1 + e \cdot x_2, 0) \quad r_2 = \max(d \cdot x_1 + f \cdot x_2, 0)$$

خروجی لایه‌ی مخفی دوم نیز به شکل زیر است.

$$s_1 = \frac{1}{1 + \exp(-a \cdot r_1)} \quad s_2 = \frac{1}{1 + \exp(-b \cdot r_2)}$$

خروجی شبکه نیز برابر $y = s_1 + s_2$ است.

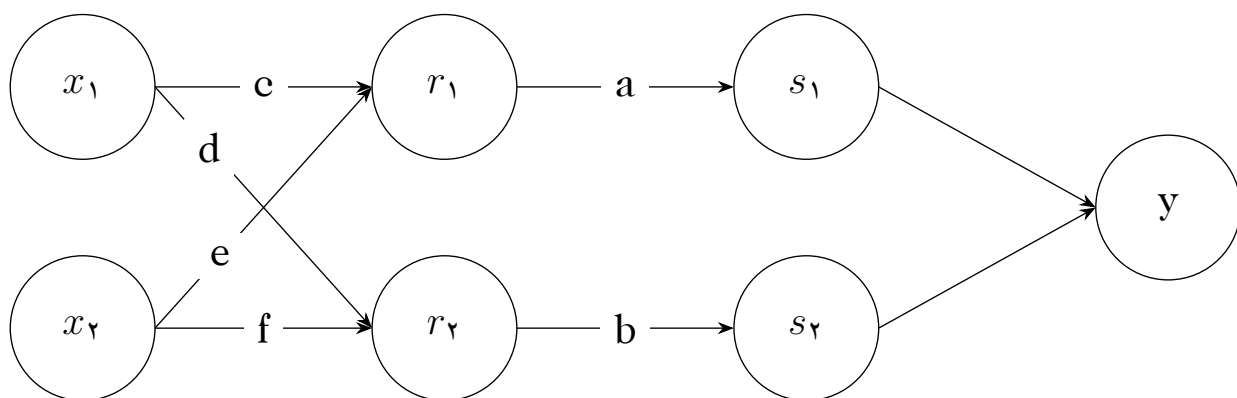
فرض کنید ورودی شبکه $x_1 = 1, x_2 = -1$ و مقدار وزن‌ها برابر

$$a = 1, b = 1, c = 4, d = 1, e = 2, f = 2$$

باشد. در این صورت خروجی تقریبی شبکه به صورت

$$r_1 = 2, r_2 = 0, s_1 = 0.9, s_2 = 0.5, y = 1.4$$

خواهد بود.



با استفاده از مقادیر تقریبی داده‌شده از الگوریتم backpropagation استفاده کنید تا مقدار مشتق‌های جزئی را حساب کنید. مقادیر را به صورت عددی به دست آورید.

(۱) مقدار $\frac{\partial y}{\partial b}$ و $\frac{\partial y}{\partial a}$ را به دست آورید.

(۲) مقدار $\frac{\partial r_1}{\partial e}$ و $\frac{\partial r_1}{\partial c}$ را به دست آورید.

(۳) مقدار $\frac{\partial r_2}{\partial f}$ و $\frac{\partial r_2}{\partial d}$ را به دست آورید.

(۴) مقدار $\frac{\partial y}{\partial r_2}$ و $\frac{\partial y}{\partial r_1}$ را به دست آورید.

(۵) با استفاده از مقادیر به دست آمده در بخش ۴ و خاصیت زنجیره‌ای مشتقات، $\frac{\partial y}{\partial e}$ و $\frac{\partial y}{\partial c}$ را به دست آورید.

(۶) با استفاده از مقادیر به دست آمده در بخش ۴ و خاصیت زنجیره‌ای مشتقات، $\frac{\partial y}{\partial f}$ و $\frac{\partial y}{\partial d}$ را به دست آورید.

(۷) تمام وزن‌های شبکه را یک مرحله با نرخ یادگیری یک بروزرسانی کنید.