۱۳۰۷
دانشگاه صنعتی خواجه نصیرالدین طوسی

دانشکده مهندسی برق _گرایش کنترل

تمرین های امتیازی ویدئو های حل تمرین

یادگیری ماشین

نگارش

فاطمه امیری

۴۰۲۰۲۴۲۴

لینک گیت هاب

استاد مربوطه

جناب آقای دکتر علیاری

تیر ماه ۱۴۰۳

بسم الله الرحمن الرحيم

سوال اول)

« فاطمه امیری »   « محبت ریاضیات »

« سؤالات امتیازی »

سؤال ۱) متغیرهای زیر را داریم:

$$y = \begin{cases} 1 & P = 0.4 \\ 2 & P = 0.1 \\ 3 & P = 0.5 \end{cases} \qquad x = \begin{cases} 1 & P = 0.8 \\ 2 & P = 0.1 \\ 3 & P = 0.1 \end{cases}$$

$$K = (w-1)^2 = \begin{cases} 0 & P_1 \\ 1 & P_0 + P_2 \end{cases} \quad , \quad \begin{cases} P = 0.8 \\ P = 0.1 \\ P = 0.1 \end{cases} , \; Z = 2x \; , \qquad w = \begin{cases} 0 & P_0 \\ 1 & P_1 \\ 2 & P_2 \end{cases}$$

و داریم $m = x + y$, $L = x - y$ $\longleftarrow$ ① m و L را محاسبه کنید؟ (مستقل اند)

② در این بخش با فرض وابسته بودن m و L ؛ محاسبه توزیع توأم m و L (برای توزیع توأم اعداد جا گذاری کنیم و مجموع آن ۱ باشد)

③ محاسبه m و L با بخش ۱

بخش ①

محاسبه M:

$$M = \begin{cases} 2 \longrightarrow P = 0.8 \times 4 = 0.32 \\ 3 \longrightarrow P = 0.8 \times 0.1 + 0.4 \times 0.1 = 0.12 \\ 4 \longrightarrow P = 0.8 \times 0.5 + 0.1 \times 0.1 + 0.1 \times 0.4 = 0.45 \\ 5 \longrightarrow P = 0.1 \times 0.1 + 0.1 \times 0.5 = 0.06 \\ 6 \longrightarrow P = 0.1 \times 0.5 = 0.05 \end{cases}$$

محاسبه L:

$$L = \begin{cases} 0 \longrightarrow P = 0.8 \times 0.4 + 0.1 \times 0.1 + 0.1 \times 0.5 = 0.38 \\ -1 \longrightarrow P = 0.8 \times 0.1 + 0.5 \times 0.1 = 0.13 \\ -2 \\ 1 \\ 2 \end{cases}$$

بخش ②

جدول توأم x و y را به صورت مقابل می‌نویسیم:

$y \backslash x$	1	2	3
1	0.32	0.08	0.4
2	0.04	0.01	0.05
3	0.04	0.01	0.05

حالا m و L و احتمال آن‌ها را حساب می‌کنیم و داریم :

x	y	m	L	احتمال
۱	۱	۲	۰	۰/۳۲
۱	۲	۳	-۱	۰/۰۸
۱	۳	۳	-۲	۰/۴
۲	۱	۴	۱	۰/۰۴
۲	۲	۳	۰	۰/۰۱
۲	۳	۴	-۱	۰/۰۵
۳	۱	۵	۲	۰/۰۴
۳	۲	۴	۱	۰/۰۱
۳	۳	۶	۰	۰/۰۵

برای جمع احتمال :

$$P = ۰/۳۲ + ۰/۰۸ + ۰/۰۴ + ۰/۴ + ۰/۰۱ + ۰/۰۴ + ۰/۰۵ + ۰/۰۱ + ۰/۰۵ = \underline{۱}$$

برابر ۱ است پس توزیع توأم صحیح است.

(۳)

m بخش ۱	L بخش ۱	احتمال بخش ۱ (برای m)	احتمال بخش ۱ (برای L)	m بخش ۲	L بخش ۲	احتمال بخش ۲
۲	۰	۰/۳۲	۰/۳۸	۲	۰	۰/۳۲
۳	-۱	۰/۱۲	۰/۱۳	۳	-۱	۰/۰۸
۴	-۲	۰/۴۵	۰/۴	۳	۱	۰/۰۴
۵	۱	۰/۰۶	۰/۰۵	۴	-۲	۰/۴
۶	۲	۰/۰۵	۰/۰۴	۴	۰	۰/۰۱
				۴	۲	۰/۰۴
				۵	-۱	۰/۰۵
				۵	۱	۰/۰۱
				۶	۰	۰/۰۵

سوال دوم)

فاطمه امیری

سوال ۲) رابطهٔ زیر را اثبات کنید؟

① $\mathbb{E}[ax+b] = a\mathbb{E}[x]+b$

② $\mathbb{E}[ag(x)+bh(x)] = a\mathbb{E}[g(x)]+b\mathbb{E}[h(x)]$

③ $\mathbb{E} = [x \pm y] = \mathbb{E}[x] \pm \mathbb{E}[y]$

④ $\mathbb{E}[xy] = \mathbb{E}[x]\mathbb{E}[y]$

اثبات ① : x: متغیر تصادفی گسسته $\longrightarrow x_1, x_2, \dots, x_n \xrightarrow[\text{متناظر}]{\text{احتمال‌های}} P_1, P_2, \dots, P_n$

$\rightarrow$ امید ریاضی : $\mathbb{E}[x] = \sum_{i=1}^{n} x_i P_i$

$ax+b$ یک تابع خطی از X است و مقادیرش به صورت ax_i+b خواهد بود. پس داریم:

$$\xrightarrow[\text{امید ریاضی}]{ax+b} \mathbb{E}[ax+b] = \sum_{i=1}^{n} \underbrace{(ax_i+b)P_i}_{(ax_iP_i+bP_i)}$$

حالا این رابطه را تفکیک می‌کنیم :

$$\mathbb{E}[ax+b] = a\underbrace{\sum_{i=1}^{n} x_iP_i}_{\text{این هم همان } \mathbb{E}[x] \text{ است}} + b\underbrace{\sum_{i=1}^{n} P_i}_{\text{مقدار این ۱ است}}$$

پس $\Longrightarrow \mathbb{E}[ax+b] = a\mathbb{E}[x]+b$

ثابت شد.

اثبات ② از خاصیت خطی بودن امید ریاضی استفاده می‌کنیم ؛ داریم :

$$\begin{cases} \Sigma P_i\, g(x_i) = \mathbb{E}[g(x)] \\ \Sigma P_i\, h(x_i) = \mathbb{E}[h(x)] \end{cases}$$

a, b ثابت اند.

$$\Rightarrow \mathbb{E}[ag(x)] + \mathbb{E}[bh(x)] \longrightarrow a\mathbb{E}[g(x)]+b\mathbb{E}[h(x)]$$

پس رابطه ثابت شد.

اثبات ③

داریم : $\begin{cases} E[x] = \sum x_i P_i \\ E[y] = \sum y_i P_i \end{cases} \Rightarrow P(x = x_i \,,\, y = y_i) = P_{ij}$ بزرگ

$$E[x \pm Y] = \sum_{ij} (x_i \pm y_i) P_{ij} \overset{\text{پس}}{=} \sum_{i,j} x_i P_{ij} \pm \sum_{i,j} y_i P_{ij} \Rightarrow \begin{cases} \sum_j P_{ij} = P_i \\ \sum_i P_{ij} = q_j \end{cases}$$

حال داریم :

$$\Rightarrow \begin{cases} \sum_{i,j} x_i P_{ij} = \sum_i x_i \left(\sum_j P_{ij} \right) = \sum_i x_i P_i = E[x] \\ \sum_{i,j} y_j P_{ij} = \sum_j y_j \left(\sum_i P_{ij} \right) = \sum_j y_j q_j = E[Y] \end{cases}$$

پس درنتیجه داریم : $E[X \pm Y] = E[x] \pm E[Y] \longrightarrow$ رابطه ثابت شد.

رابطه ④

وقتی x و y مستقل اند از تعریف امید ریاضی و استقلال متغیرهای تصادفی استفاده می کنیم :

$$\begin{cases} E[x] = \sum x_i P_i \\ E[Y] = \sum_j y_j P_j \end{cases} \Rightarrow E[xY] = \sum_i \sum_j (x_i y_j) P(X = x_i \,, Y = y_j) =$$

$$\sum_i x_i P_i \sum y_j P_j = \left(\sum_i x_i P_i \right) \left(\sum_j y_j P_j \right)$$

که داریم $E[xy] = E[x] \cdot E[y] \longrightarrow$ رابطه ثابت شد.

سوال سوم)

فاطمه امیری

«مبحث درخت تصمیم»

برای x_2 و x_4 ، IG را محاسبه کنید؟

$x_2 \rightarrow E = 0.59$ و $x_4 \rightarrow E = 0$

$$IG = E(\text{parent}) - \sum w_i E_i (\text{child})$$

$$\begin{cases} IG_{x_2} = 0.69 - 0.56 = \underline{\underline{0.13}} \\ IG_{x_4} = 0.69 - 0 = \underline{\underline{0.69}} \end{cases}$$

سوال چهارم)

برای جلوگیری از اورفیت در درخت تصمیم ، چی کار میشه کرد ؟

درخت تصمیم یکی از مدل‌های پرکاربرد در یادگیری ماشین است که با استفاده از ساختار درختی، تصمیم‌گیری و پیش‌بینی را انجام می‌دهد. اما یکی از مشکلاتی که ممکن است در این مدل رخ دهد، **بیش‌برازش (overfitting)** است. بیش‌برازش زمانی رخ می‌دهد که مدل به جای یادگیری الگوهای کلی داده‌ها، به یادگیری جزئیات و نوسانات تصادفی داده‌های آموزشی می‌پردازد. در نتیجه، مدل نمی‌تواند به خوبی داده‌های جدید و نادیده را پیش‌بینی کند.

روش‌های جلوگیری از بیش‌برازش در درخت تصمیم

۱. **هرس درخت (Pruning):** هرس درخت به دو صورت می‌تواند انجام شود: پیش از ساخت درخت و پس از ساخت درخت.
   - **هرس پیش از ساخت درخت (Pre-pruning):** در این روش، ساخت درخت در مرحله‌ای خاص متوقف می‌شود تا از پیچیدگی بیش از حد جلوگیری شود. برخی از معیارهای مورد استفاده برای توقف ساخت درخت شامل موارد زیر است:
     - **حداکثر عمق درخت (max_depth):** با محدود کردن عمق درخت، از ایجاد گره‌های بیش از حد جلوگیری می‌شود.
     - **حداقل تعداد نمونه‌ها در هر گره (min_samples_split):** تعیین می‌کند که یک گره تنها در صورتی تقسیم شود که تعداد نمونه‌های موجود در آن بیشتر از مقدار مشخصی باشد.
     - **حداکثر تعداد گره‌های برگ (max_leaf_nodes):** محدود کردن تعداد گره‌های برگ به یک مقدار مشخص، از ایجاد گره‌های بیش از حد جلوگیری می‌کند.
   - **هرس پس از ساخت درخت (Post-pruning):** در این روش، پس از ساخت درخت، برگ‌هایی که اهمیت کمی دارند به صورت تدریجی حذف می‌شوند. این کار با استفاده از پارامتر **ccp_alpha**در Scikit-learn انجام می‌شود که تعادل بین بایاس و واریانس را تنظیم می‌کند.

۲. **تنظیم حداقل تعداد نمونه‌های لازم برای تقسیم یک گره:** با استفاده از پارامتر **minimum_samples_split**، می‌توان تعیین کرد که یک گره تنها در صورتی تقسیم شود که تعداد

نمونه‌های موجود در آن بیشتر از مقدار مشخصی باشد. این کار از تقسیم گره‌های با داده‌های کم و نامناسب جلوگیری می‌کند و قابلیت تعمیم مدل را افزایش می‌دهد.

۳. **استفاده از اعتبارسنجی متقاطع (Cross-validation):** اعتبارسنجی متقاطع شامل تقسیم داده‌ها به چندین بخش و استفاده از هر بخش به عنوان داده‌های تست به صورت متناوب است. این روش باعث می‌شود که مدل به طور مکرر با داده‌های جدید تست شود و از دقت بالای آن در پیش‌بینی داده‌های جدید اطمینان حاصل شود. این روش به طور موثر به جلوگیری از بیش‌برازش کمک می‌کند.

همچنین مقاله ی Classification and Regression Trees (CART) توسط Breiman, Friedman, Olshen, و Stone (1984) یکی از منابع اصلی و کلاسیک در زمینه درخت‌های تصمیم است که به طور جامع به معرفی و توضیح مفاهیم اساسی درخت‌های تصمیم، از جمله ساختار، مزایا، معایب و کاربردهای آنها می‌پردازد. نویسندگان در این اثر، تکنیک‌های مختلف برای بهینه‌سازی درخت‌های تصمیم را بررسی می‌کنند و به طور ویژه به هرس درخت به عنوان یکی از راهکارهای موثر برای جلوگیری از بیش‌برازش توجه می‌کنند.

پس با بهره‌گیری از تکنیک‌هایی که بیان کردیم، می‌توان از بیش‌برازش در درخت تصمیم جلوگیری کرده و مدلی با قابلیت تعمیم بالا و دقت مناسب ایجاد کرد. این مدل‌ها قادر خواهند بود داده‌های جدید را به خوبی پیش‌بینی کنند و از پیچیدگی بی‌مورد و بیش از حد جلوگیری کنند.

سوال پنجم)

در معیار آماری Kullback-Leibler divergence ، آیا نحوه توزیع $|X-Y|$ با نحوه توزیع $|Y-X|$ برابر است؟ آیا رابطه زیر برقرار است؟ توضیح دهید :

$$D_{KL}(P||Q) = \sum_x P(x) \log\left(\frac{P(x)}{Q(x)}\right) =^?= D_{KL}(Q||P) = \sum_x Q(x) \log\left(\frac{Q(x)}{P(x)}\right)$$

Kullback- Leibler drivergence یک معیار آماری است برای اندازه‌گیری تفاوت بین ۲ توزیع احتمالی P و Q و کمک به سنجش میزان اطلاعات از دست رفته زمانی که از یک توزیع به جای دیگری استفاده می‌کنیم، است.

K_L divergence بین توزیع P و Q $\longrightarrow D_{KL}(P||Q) = \sum_x P(x) \log\left(\frac{P(x)}{Q(x)}\right)$

حالت پیوسته انتگرالی $\longrightarrow D_{KL}(P||Q) = \int_{-\infty}^{\infty} P(x) \log\left(\frac{P(x)}{Q(x)}\right) dx$

ویژگی مهم KL divergence نامتقارن بودن آن است یعنی به طور کلی : $D_{KL}(P||Q) \neq D_{KL}(Q||P)$

این عدم تقارن نشان می دهد که میزان اطلاعات از دست رفته بسته به اینکه کدام توزیع را به عنوان اصلی و کدام را به عنوان جایگزین در نظر بگیریم، متفاوت است. KL divergence همیشه غیرمنفی است و فقط زمانی که P و Q یکسان باشند مقدار آن صفر می‌شود. همچنین درحالی که $|x-y|$ $|x-y|$ $|x-y|$ و $|y-x|$ $|y-x|$ $|y-x|$ $|y-x|$ همیشه برابرند، KL divergence به ترتیب توزیع ها حساس است و مقادیر متفاوتی می‌تواند داشته باشد.

درواقع اگر x و y دو متغیر تصادفی باشند، $|x-y|$ و $|y-x|$ همیشه برابرند چون قدر مطلق اند، ولی KL divergence $D_{KL}(P||Q)$ و $D_{KL}(Q||P)$ برابر نیستند (همان طور که گفتم به ترتیب توزیع ها وابسته است.)

پس KL divergence یک معیار نامتقارن است.