به نام خدا

نام و نام خانوادگی: ریحانه صحراکار

شماره دانشجویی: ٤٠٤١٥٠٠١٠٠٦

پاسخ‌های تشریحی به سوالات کتاب "Pattern Recognition and Machine Learning"

سوال ۱

الف) محاسبه امید ریاضی E[X]

برای محاسبه امید ریاضی یک متغیر تصادفی پیوسته، از انتگرال زیر استفاده می‌کنیم:

$$E[X] = \int_0^1 x \cdot p(x)\, dx$$

با جایگذاری تابع چگالی احتمال داده شده:

$$E[X] = \int_0^1 x \cdot (3x^2)\, dx = \int_0^1 3x^3\, dx$$

انتگرال را محاسبه می‌کنیم:

$$E[X] = 3 \cdot [x^4/4]_0^1 = 3 \cdot (1/4 - 0) = 3/4 = 0.75$$

تحلیل: امید ریاضی یا میانگین این توزیع ۰٫۷۵ است که نشان می‌دهد مقادیر متغیر تصادفی تمایل دارند به سمت ۱ گرایش داشته باشند، زیرا تابع چگالی احتمال با افزایش x افزایش می‌یابد.

ب) محاسبه واریانس Var[X]

واریانس از رابطه زیر محاسبه می‌شود:

$$Var[X] = E[X^2] - (E[X])^2$$

ابتدا $E[X^2]$ را محاسبه می‌کنیم:

$$E[X^2] = \int_0^1 x^2 \cdot (3x^2)\, dx = \int_0^1 3x^4\, dx = 3 \cdot [x^5/5]_0^1 = 3/5 = 0.6$$

حال واریانس را محاسبه می‌کنیم:

$$Var[X] = E[X^2] - (E[X])^2 = 0.6 - (0.75)^2 = 0.6 - 0.5625 = 0.0375$$

یا به صورت کسری:

$$Var[X] = 3/5 - (3/4)^2 = 3/5 - 9/16 = (48/80) - (45/80) = 3/80 = 0.0375$$

تحلیل: واریانس ۰٫۰۳۷۵ نسبتاً کوچک است که نشان می‌دهد مقادیر متغیر تصادفی حول میانگین ۰٫۷۵ متمرکز هستند. این با شکل تابع چگالی احتمال که یک تابع افزایشی است همخوانی دارد.

ج) تحلیل تغییر تابع چگالی احتمال به $p(x) = 2x^3$

ابتدا باید ضریب نرمال‌سازی را محاسبه کنیم. برای یک تابع چگالی احتمال معتبر باید داشته باشیم:

$$\int_0^1 p(x)\,dx = 1$$

برای تابع جدید:

$$\int_0^1 2x^3\,dx = 2 \cdot [x^4/4]_0^1 = 2 \cdot (1/4) = 1/2 \neq 1$$

بنابراین برای نرمال‌سازی، تابع چگالی احتمال صحیح باید باشد:

$p(x) = 4x^3$ (چون $2/(1/2) = 4$)

محاسبات برای تابع جدید:

۱. امید ریاضی:

$$E[X] = \int_0^1 x \cdot (4x^3)\,dx = \int_0^1 4x^4\,dx = 4 \cdot [x^5/5]_0^1 = 4/5 = 0.8$$

۲. $E[X^2]$:

$$E[X^2] = \int_0^1 x^2 \cdot (4x^3)\,dx = \int_0^1 4x^5\,dx = 4 \cdot [x^6/6]_0^1 = 4/6 = 2/3 \approx 0.6667$$

۳. واریانس:

$$Var[X] = E[X^2] - (E[X])^2 = 2/3 - (4/5)^2 = 2/3 - 16/25 = (50/75) - (48/75) = 2/75 \approx 0.02667$$

تحلیل مقایسه‌ای:

- امید ریاضی: از ۰,۷۵ به ۰,۸ افزایش یافته است.
- واریانس: از ۰,۰۳۷۵ به ۰,۰۲۶۶۷ کاهش یافته است.

دلیل تغییرات:

تابع جدید $p(x) = 4x^3$ نسبت به تابع قبلی $p(x) = 3x^2$ برای مقادیر x نزدیک به ۱ وزن بیشتری اختصاص می‌دهد (چون x^3 سریع‌تر از x^2 رشد می‌کند). این باعث می‌شود:

۱. میانگین به سمت ۱ جابجا شود (افزایش امید ریاضی)

۲. توزیع حول میانگین جدید متمرکزتر شود (کاهش واریانس)

به عبارت دیگر، توزیع جدید "سنگین‌تر" در دم بالایی است که منجر به میانگین بالاتر و پراکندگی کمتر می‌شود.

سوال ۲

الف) محاسبه آنتروپی توزیع $p = (0.3, 0.2, 0.4, 0.1)$

آنتروپی شانون با فرمول زیر محاسبه می‌شود:

$$H(p) = -\sum p_i \log_2(p_i)$$

برای توزیع داده شده:

$$H(p) = -[0.3\cdot\log_2(0.3) + 0.2\cdot\log_2(0.2) + 0.4\cdot\log_2(0.4) + 0.1\cdot\log_2(0.1)]$$

محاسبه مقادیر لگاریتم:

$\log_2(0.3) \approx -1.7370$

$\log_2(0.2) \approx -2.3219$

$\log_2(0.4) \approx -1.3219$

$\log_2(0.1) \approx -3.3219$

حال مقادیر را جایگزین می‌کنیم:

$$H(p) = -[0.3\times(-1.7370) + 0.2\times(-2.3219) + 0.4\times(-1.3219) + 0.1\times(-3.3219)]$$
$$= -[(-0.5211) + (-0.4644) + (-0.5288) + (-0.3322)]$$
$$= -[-1.8465] = 1.8465 \text{ بیت}$$

تحلیل: آنتروپی تقریباً ۱٫۸۵ بیت نشان می‌دهد که متوسط اطلاعات مورد نیاز برای توصیف نتیجه این متغیر تصادفی حدود ۱٫۸۵ بیت است.

ب) محاسبه آنتروپی توزیع یکنواخت و مقایسه

برای توزیع یکنواخت با ۴ حالت ممکن:

$p = (0.25, 0.25, 0.25, 0.25)$

$H_uniform = -4 \times [0.25 \times \log_2(0.25)]$

از آنجایی که: $\log_2(0.25) = \log_2(1/4) = -2$

$H_uniform = -4 \times [0.25 \times (-2)] = -4 \times [-0.5] = 2$ بیت

مقایسه:

- آنتروپی توزیع اولیه: ۱,۸٤٦٥ بیت
- آنتروپی توزیع یکنواخت: ۲ بیت
- تفاوت: ۰,۱٥۳٥ بیت

چرا آنتروپی برای توزیع یکنواخت بیشتر است؟

آنتروپی معیاری از عدم قطعیت یا بی‌نظمی در یک توزیع احتمال است. در توزیع یکنواخت، همه حالات به یک اندازه محتمل هستند، بنابراین:

- بیشترین عدم قطعیت را داریم

- بیشترین غافلگیری را هنگام مشاهده نتیجه تجربه می‌کنیم

- بیشترین اطلاعات را برای توصیف سیستم نیاز داریم

در توزیع اولیه، برخی حالات (مثلاً حالت سوم با احتمال ۰,٤) محتمل‌تر هستند، بنابراین:

- عدم قطعیت کمتر است

- تا حدی می‌توانیم نتیجه را پیش‌بینی کنیم

- اطلاعات کمتری برای توصیف سیستم نیاز داریم

از دیدگاه ریاضی، آنتروپی برای توزیع یکنواخت بیشینه است زیرا تابع $-p \log p$ یک تابع مقعر است و تحت محدودیت $\sum p_i = ۱$، در نقطه‌ای که همه p_i برابر باشند بیشینه می‌شود.

ج) تحلیل نقش آنتروپی در یادگیری ماشین و تصمیم‌گیری

اهمیت آنتروپی در یادگیری ماشین:

۱. معیار عدم قطعیت: آنتروپی میزان عدم اطمینان را در پیش‌بینی‌های مدل اندازه‌گیری می‌کند. مدلی با آنتروپی پایین، اطمینان بیشتری در پیش‌بینی‌های خود دارد.

۲. انتخاب ویژگی در درخت‌های تصمیم: در الگوریتم‌هایی مانند ID3 و C4.5، از «افزایش اطلاعات» (Information Gain) برای انتخاب ویژگی‌ها استفاده می‌شود که مبتنی بر کاهش آنتروپی است:

Information Gain = آنتروپی قبل از تقسیم - آنتروپی بعد از تقسیم

- ویژگی‌ای انتخاب می‌شود که بیشترین کاهش آنتروپی (بیشترین افزایش اطلاعات) را ایجاد کند.

۳. نظریه تصمیم بیزی: در تصمیم‌گیری تحت عدم قطعیت، آنتروپی می‌تواند به عنوان معیاری برای ارزیابی توزیع پسین استفاده شود. توزیع‌های با آنتروپی پایین ترجیح داده می‌شوند زیرا نمایانگر اطمینان بیشتر هستند.

٤. منظم‌سازی (Regularization): در برخی مدل‌ها، افزودن جمله‌ای بر پایه آنتروپی به تابع هزینه، از بیش‌برازش جلوگیری می‌کند.

چرا اهمیت دارد که آنتروپی به حداقل برسد؟

۱. پیش‌بینی‌های قطعی‌تر: در بسیاری از کاربردهای عملی (مانند تشخیص پزشکی، سیستم‌های توصیه‌گر، تشخیص تقلب)، نیاز به پیش‌بینی‌های قطعی داریم. مدل‌هایی با آنتروپی پایین، تصمیم‌های واضح‌تری می‌گیرند.

۲. کارایی محاسباتی: مدل‌هایی با آنتروپی پایین معمولاً ساده‌تر هستند و نیاز به منابع محاسباتی کمتری دارند.

۳. قابلیت تفسیر: مدل‌هایی که آنتروپی را کمینه می‌کنند (مانند درخت‌های تصمیم با عمق محدود)، معمولاً قابل تفسیرتر هستند.

٤. بهبود عملکرد: در بسیاری از سناریوها، کمینه کردن آنتروپی منجر به بهبود معیارهای ارزیابی مانند دقت، صحت و بازیابی می‌شود.

تعادل بین آنتروپی و پیچیدگی: گاهی کمینه کردن بیش از حد آنتروپی می‌تواند منجر به مدل‌های بسیار ساده‌ای شود که نمی‌توانند پیچیدگی داده‌ها را به خوبی مدل کنند (تحت‌برازش).

نتیجه‌گیری: آنتروپی یک مفهوم اساسی در یادگیری ماشین است که هم به عنوان معیار ارزیابی و هم به عنوان بخشی از فرآیند یادگیری استفاده می‌شود. درک عمیق آن برای طراحی مدل‌های کارآمد و تفسیر نتایج آن‌ها ضروری است. هدف معمولاً یافتن تعادل بهینه بین کاهش آنتروپی (برای قطعیت بیشتر) و حفظ انعطاف‌پذیری مدل (برای تطبیق با داده‌های جدید) است.