#خوانش_گروهی_با_فیلاگر

مطالعه کتاب و مقالات هوش‌مصنوعی به صورت گروهی در فیلاگر

رقیه فرجی

کارشناسی ارشد فیزیک اتمی

فیلاگر|جامعه هوش‌مصنوعی ایران

فصل سوم

تئوری احتمالات و اطلاعات

- ✍️ مقدمه‌ای بر احتمال
- ✍️ متغیرهای تصادفی
- ✍️ تابع توزیع احتمال
- ✍️ Marginal Probability
- ✍️ احتمال شرطی و قاعده زنجیری
- ✍️ مقداری انتظاری، واریانس و کواریانس

مقدمه بر احتمال

تئوری احتمال نوعی از چهارچوب ریاضیاتی برای نمایش پدیده‌های غیرقطعی ارائه می‌دهد.

به دو دلیل عمده نظریه احتمالات در مباحث هوش مصنوعی حائز اهمیت است:

- قوانین احتمالات قدرت و نحوه استدلال را به مدلAI می‌دهد.
- برای تجزیه و تحلیل رفتار سیستم AI نیز می‌توان از نظریه احتمالات استفاده کرد

54

در حالی که نظریه احتمال به ما امکان می دهد در صورت عدم وجود قطعیت نمایشی برای رخداد پدیده‌ها ارائه دهیم ، تئوری اطلاعات به ما اجازه می دهد تا مقدار عدم قطعیت را در یک توزیع احتمال کمی کنیم.

مقدمه بر احتمال

دلیل‌های بسیاری برای ورود عدم قطعیت به مباحث یادگیری‌ماشین وجود دارد که در زیر به برخی از آن‌ها می‌پردازیم:

- ماهیت تصادفی بودن: ماهیت احتمالاتی که ناشی از مکانیک کوانتومی می‌باشد
- اندازه‌گیری ناکامل:هر پدیده قابل اندازه‌گیری هم اگر همه متغیرهایش را اندازه‌گیری نکنیم می‌تواند یک پدیده تصادفی تلقی شود
- مدلسازی ناقص: در هنگام مدلسازی برخی اطلاعات از بین می‌رود که این می‌تواند باعث ایجاد عدم قطعیت شود

54
55

مقدمه بر احتمال

- گاهی‌اوقات استفاده از یک مدل با uncertainty ساده بسیار راه‌گشاتر از مدل دقیق اما پیچیده است
- تئوری احتمالات اساساً برای تعداد تکرارهای وقایع رخدهنده بوجود آمده است
- در وقایع تکرار شونده هنگامی که می‌گوییم احتمال یک رخداد p می‌باشد بدین معناست که اگر به تعداد دفعات زیادی این پدیده اندازه‌گیری شود نسبت p تا از آن پیشامد موردنظر اتفاق می‌افتد . به این نوع پدیده‌های تکرار شونده frequentist probability می‌گویند
- برخی موارد وجود دارند که قابلیت تکرار شدن ندارند مثلا هنگامیکه یک پزشک احتمال ابتلا یک فرد به آنفولانزا را ۴۰درصد بیان می‌کند . این نوع احتمال با درجه اطمینان(degree of belief) بیان می‌شود و به این نوع پدیده Bayesian probability گفته می‌شود

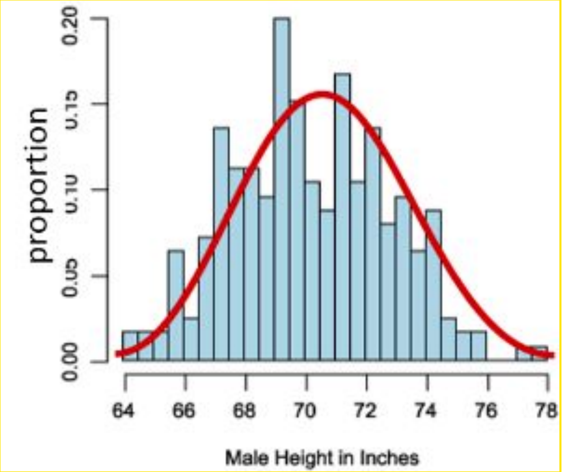
54
55

توزیع احتمال(Probability Distributions)

توصیفی از اینکه یک متغیر تصادفی یا مجموعه‌ای از متغیرهای تصادفی در هر حالت چه احتمالی برای رخدادن دارند

تابع توزیع احتمال می‌تواند بر اساس نوع متغیرهای آن می‌تواند به دو دسته تقسیم شود:

- متغیرهای گسسته: Probability Mass Function
- متغیرهای پیوسته: Probability Density Function



56

متغیرهای گسسته و Probability Mass Functions

توزیع احتمال بر روی مجموعه متغیرهای گسسته با تابع PMF نشان داده می‌شود

PMF هر متغیر تصادفی را به میزان احتمال رخدادن آن نگاشت می‌کند

P(x)=1 **certain**

P(x)=0 **impossible**

PMF از سه قاعده زیر پیروی می‌کند:

- دامنه اعدادی که در تابع PMF قرار می‌گیرند باید از مجموعه x باشد
- $\forall x \in \mathrm{x}, 0 \leq P(x) \leq 1.$
- $\sum_{x \in \mathrm{x}} P(x) = 1.$

57

متغیرهای گسسته و Probability Mass Functions

بعنوان مثال فرض کنید مجموعه‌ای از اعداد گسسته k عضوی داریم تابع PMF آن به شکل مقابل تعریف می‌شود

$$\sum_i P(\mathrm{x} = x_i) = \sum_i \frac{1}{k} = \frac{k}{k} = 1,$$

$$P(\mathrm{x} = x_i) = \frac{1}{k}$$

PMF می‌تواند همزمان روی چند متغیر عمل کند به‌عنوان مثال می‌تواند تابعی از x و y باشد در این صورت توزیع آن joint probability distribution نامیده می‌شود

متغیرهای پیوسته و Probability Density Functions

هنگامی که متغیرهای تصادفی پیوسته باشند توزیع احتمال آن‌ها بصورت PDF تعریف می‌شود

در متغیرهای پیوسته تابع PDF بصورت صریح احتمال در یک نقطه x را نمی‌دهد بلکه در بازه‌ی بسیار کوچک δx احتمال رخدادی به‌صورت $p(x)\delta x$ می‌دهد

- دامنه اعدادی که در تابع PDF قرار می‌گیرند باید از بازه مجاز x باشد
- $\forall x \in \mathrm{x}, p(x) \geq 0.$
- $\int p(x)dx = 1.$

لزوما $p(x) \leq 1.$ می‌تواند برقرار نباشد

Marginal Probability

گاهی اوقات ما توزیع احتمال را روی مجموعه ای از متغیرها می دانیم و می خواهیم توزیع احتمال را فقط روی زیرمجموعه ای از آنها بدانیم

متغیرهای گسسته

$$\forall x \in \mathrm{x}, P(\mathrm{x}=x)=\sum_{y} P(\mathrm{x}=x, \mathrm{y}=y).$$

متغیرهای پیوسته

$$p(x)=\int p(x, y) dy.$$

P(Wed.)

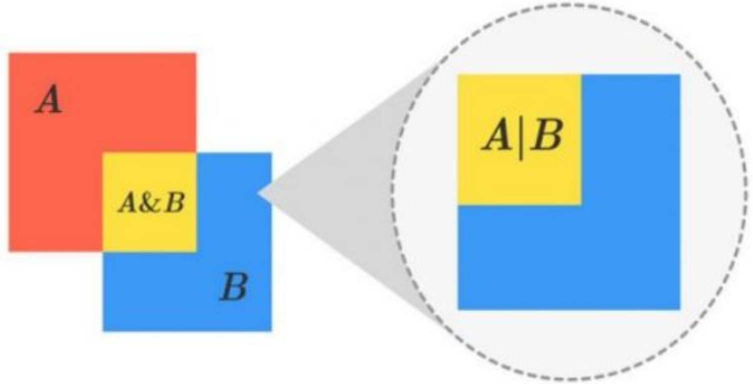
$$= P(\text{Jan. and Wed.}) + P(\text{Not Jan. and Wed.}) = \frac{5}{365} + \frac{48}{365} = \frac{53}{365}$$

	Jan.	Not Jan.	Total
Wed.	5	48	52
Not Wed.	27	286	313
Total	31	334	365

Conditional Probability

در بسیاری از موارد با توجه به اینکه رخدادی اتفاق افتاده است به دنبال یافتن احتمال پدیده‌ای هستیم اینگونه احتمال‌ها را احتمال شرطی می‌نامیم

$$P(\mathrm{y} = y \mid \mathrm{x} = x) = \frac{P(\mathrm{y} = y, \mathrm{x} = x)}{P(\mathrm{x} = x)}.$$



59

The Chain Rule of Conditional Probabilities

برای محاسبه joint probability که شامل چند متغیر تصادفی است می‌توان آن را به ضرب احتمالات شرطی تبدیل کرد

$$P(\mathrm{x}^{(1)}, \ldots, \mathrm{x}^{(n)}) = P(\mathrm{x}^{(1)}) \Pi_{i=2}^{n} P(\mathrm{x}^{(i)} \mid \mathrm{x}^{(1)}, \ldots, \mathrm{x}^{(i-1)}).$$

60

مثال

$$P(\mathrm{a}, \mathrm{b}, \mathrm{c}) = P(\mathrm{a} \mid \mathrm{b}, \mathrm{c}) P(\mathrm{b}, \mathrm{c}) \quad 1$$

$$2 \quad P(\mathrm{b}, \mathrm{c}) = P(\mathrm{b} \mid \mathrm{c}) P(\mathrm{c})$$

$$P(\mathrm{a}, \mathrm{b}, \mathrm{c}) = P(\mathrm{a} \mid \mathrm{b}, \mathrm{c}) P(\mathrm{b} \mid \mathrm{c}) P(\mathrm{c}). \quad 3$$

Independence and Conditional Independence

دو متغیر تصادفی x و y را مستقل می‌گوییم هرگاه joint probability آن‌ها بصورت حاصل‌ضرب تک‌تک آن‌ها شود.

$$\forall x \in \mathrm{x}, y \in \mathrm{y},\ p(\mathrm{x} = x, \mathrm{y} = y) = p(\mathrm{x} = x)p(\mathrm{y} = y).$$

60

دو متغیر تصادفی x و y را مستقل شرطی می‌گوییم هرگاه به‌ازای متغیر سوم z رابطه زیر برقرار باشد:

$$\forall x \in \mathrm{x}, y \in \mathrm{y}, z \in \mathrm{z},\ p(\mathrm{x} = x, \mathrm{y} = y \mid \mathrm{z} = z) = p(\mathrm{x} = x \mid \mathrm{z} = z)p(\mathrm{y} = y \mid \mathrm{z} = z).$$

Expectation, Variance and Covariance

مقدار انتظاری (expectation value) برای تابع f به‌صورت مقابل می‌باشد:

$$\mathbb{E}_{\mathrm{x}\sim P}[f(x)] = \sum_x P(x)f(x),$$

$$\mathbb{E}_{\mathrm{x}\sim p}[f(x)] = \int p(x)f(x)dx.$$

61

$$\mathbb{E}_{\mathrm{x}}[\alpha f(x) + \beta g(x)] = \alpha\mathbb{E}_{\mathrm{x}}[f(x)] + \beta\mathbb{E}_{\mathrm{x}}[g(x)]$$

خاصیت خطی بودن مقدار انظاری

Expectation, Variance and Covariance

واریانس معیاری از میزان افت و خیز یک تابع حول مقدار انتظاری می‌باشد

$$\mathrm{Var}(f(x)) = \mathbb{E}\left[(f(x) - \mathbb{E}[f(x)])^2\right].$$

- **Covariance** دو متغیر تصادفی معیاری از میزان وابستگی خطی آن‌ها در تغییراتشان است

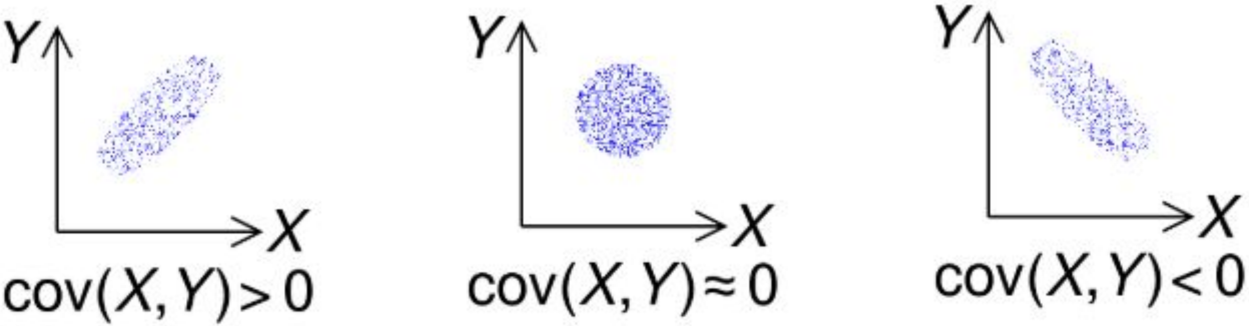
- مقادیر بالای مطلق کوواریانس به این معنی است که مقادیر متغیرهای تصادفی بسیار تغییر می کنند و هر دو در یک زمان از میانگین مربوطه فاصله می‌گیرند.

- مقدار منفی برای **Covariance** بدین معنی است که یکی از متغیرها مقدار بیشتری به خود می‌گیرد و متغیر دیگری میزان کمتری به خود می‌گیرد.

$$\mathrm{Cov}(f(x), g(y)) = \mathbb{E}\left[(f(x) - \mathbb{E}[f(x)])(g(y) - \mathbb{E}[g(y)])\right].$$

61

Expectation, Variance and Covariance



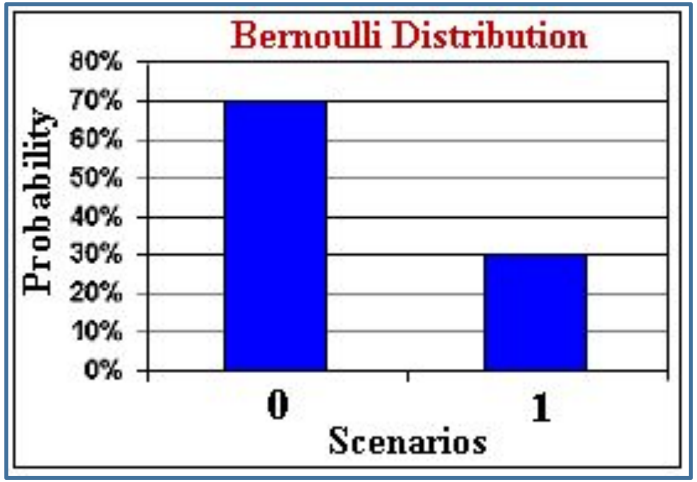
61

$$Correlation = \frac{Cov\ (x, y)}{\sigma x * \sigma y}$$

Variable	A	B	C	D
A	150	−90	100	70
B	−90	210	45	30
C	100	45	300	−85
D	70	30	−85	240

Bernoulli Distribution

توزیع برنولی توزیعی است که خروجی آن تنها دو حالت می‌تواند داشته باشد



$$P(\mathrm{x}=1)=\phi$$

$$P(\mathrm{x}=0)=1-\phi$$

$$P(\mathrm{x}=x)=\phi^{x}(1-\phi)^{1-x}$$

$$\mathbb{E}_{\mathrm{x}}[\mathrm{x}]=\phi$$

$$\mathrm{Var}_{\mathrm{x}}(\mathrm{x})=\phi(1-\phi)$$

62

توزیع مالتی‌نولی توزیعی بر روی k حالت گسسته است.(Multinoulli Distribution)

Gaussian Distribution

یکی از توزیع‌های معمولی که برای متغیرهای پیوسته بکار می‌رود توزیع نرمال یا همان Gaussian Distribution می‌باشد.

$$\mathbb{E}[\mathrm{x}] = \mu.$$

$$\mathcal{N}(x;\mu,\sigma^2) = \sqrt{\frac{1}{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right).$$

$$\boxed{\mathcal{N}(x;\mu,\beta^{-1}) = \sqrt{\frac{\beta}{2\pi}} \exp\left(-\frac{1}{2}\beta(x-\mu)^2\right).}$$

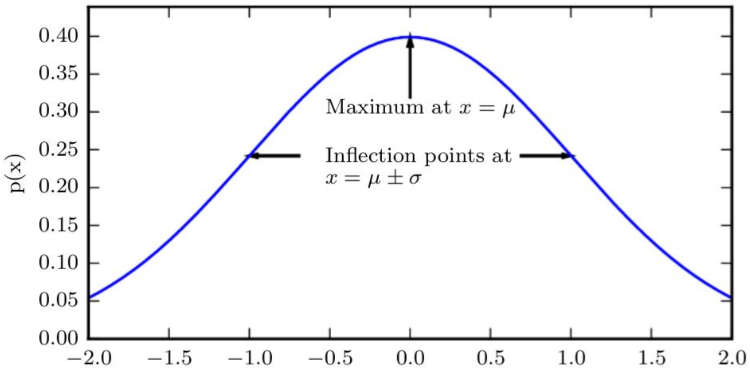
63

Gaussian Distribution

زمانیکه هیچ دانش قبلی راجعبه توزیع یک جامعه آماری نداریم توزیع نرمال به دو دلیل بهترین انتخاب است:

- بسیاری از توریع‌هایی که بررسی می‌کنیم نزدیک به توزیع نرمال هستند.
- در بین توزیع‌های موجود بیشتر عدم اطمینان را به متغیر ها می‌دهد

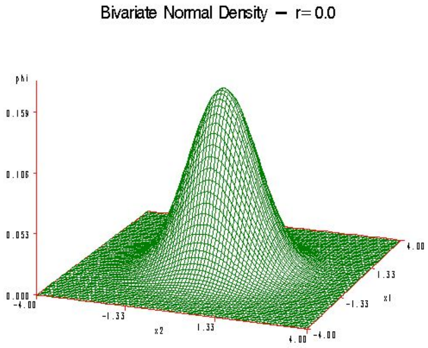
64



Multivariate Normal Distribution

$$\mathcal{N}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \sqrt{\frac{1}{(2\pi)^n \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

65



$$\mathcal{N}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\beta}^{-1}) = \sqrt{\frac{\det(\boldsymbol{\beta})}{(2\pi)^n}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\beta}(\boldsymbol{x}-\boldsymbol{\mu})\right).$$

Exponential and Laplace Distributions

گاهی اوقات در مباحث یادگیری عمیق در x=0 احتمال دارای sharp point است در این مواقع از توزیع نمایی استفاده می‌کنیم

$$p(x;\lambda) = \lambda \mathbf{1}_{x\geq 0} \exp(-\lambda x)\,.$$

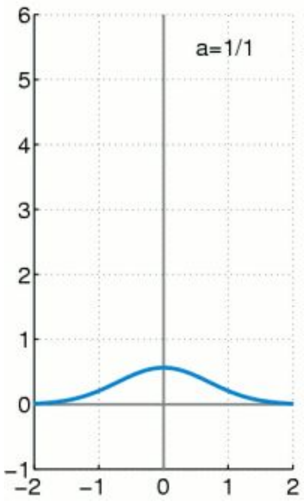
$$\text{Laplace}(x;\mu,\gamma) = \frac{1}{2\gamma}\exp\left(-\frac{|x-\mu|}{\gamma}\right).$$

Dirac Distribution

$$p(x) = \delta(x - \mu).$$

گاهی‌اوقات تمام وزن‌های احتمالی در یک توزیع حول یک تک‌نقطه است .
در این حالت تابع دلتای دیراک گزینه مناسبی است.

میزان احتمال تابع دیراک در مقدار یک بوده ودر بقیه جاها صفر می‌شود



66

سپاس از همراهی شما

فیلاگر | جامعه هوش‌مصنوعی ایران