علی عسگری - 40032223 - پایگاه داده پیشرفته

این فایل حاوی توضیح عبارات ریاضی میباشد و فایل دیگری حاوی کد ها و خروجی و توضیح کد ها ضمیمه شده است. فرمت آن فایل به شکل html است با کلیک بر روی آن ، آن را در یک مرورگر باز نمایید.

1 - میانگین :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + \cdots + x_n)$$

2 - ماتریس کوواریانس ماتریسی است که اعضای آن همبستگی بین پارامترهای مختلف سیستم را نشان می‌دهند.

اگر اعضای بردار X، متغیرهای تصادفی ما باشند:

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$$

ماتریس کوواریانس، ماتریسی است که درایه‌های آن به طریق ذیل بدست می‌آیند:

$$\Sigma_{ij} = \mathrm{cov}(X_i, X_j) = \mathrm{E}[\,(X_i - \mu_i)(X_j - \mu_j)\,]$$

که ماتریس زیر به دست می‌آید:

$$\Sigma = \begin{bmatrix} \mathrm{E}[(X_1-\mu_1)(X_1-\mu_1)] & \mathrm{E}[(X_1-\mu_1)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_1-\mu_1)(X_n-\mu_n)] \\ \mathrm{E}[(X_2-\mu_2)(X_1-\mu_1)] & \mathrm{E}[(X_2-\mu_2)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_2-\mu_2)(X_n-\mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}[(X_n-\mu_n)(X_1-\mu_1)] & \mathrm{E}[(X_n-\mu_n)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_n-\mu_n)(X_n-\mu_n)] \end{bmatrix}.$$

که در آن E امید ریاضی می باشد.

امید ریاضی از یک متغیر تصادفی، مقدارِ میانگینِ تعداد دفعاتِ مشاهده شدهٔ یک وضعیت است. به عنوان مثال، در پرتاب یک سکه، برای بدست آوردن احتمال مشاهدهٔ هر سمت از یک سکه (شیر یا خط)، می‌توان این کار را به دفعات زیاد انجام داد. اکنون میانگین تعداد دفعاتِ مشاهدهٔ هرکدام از حالت‌ها (شیر یا خط)، برابر است با امید ریاضی به‌طور مثال برای تاس داریم : یعنی اگر بی‌نهایت بار تاس را پرت کنیم، مقدار میانگین بدست آمده به سمت عدد ۳,۵ میل خواهد کرد.

$$\mathbb{E}[X] = \frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \frac{1}{6} \times 3 + \frac{1}{6} \times 4 + \frac{1}{6} \times 5 + \frac{1}{6} \times 6 = 3.5$$

به عبارت دیگر اما میتوان نوشت ، کوواریانس دو بردار x و y به صورت زیر تعریف می شود:

$$Cov(x,y) = \frac{1}{n-1}\sum_{i=0}^{n}(x_i - \bar{x})(x_i - \bar{x})^T$$

$$Cov(X) = \begin{bmatrix} cov(x_1,x_1) & \cdots & cov(x_1,x_n) \\ \vdots & \ddots & \vdots \\ cov(x_n,x_1) & \cdots & cov(x_n,x_n) \end{bmatrix}$$

به یک مثال توجه کنید :

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

$$\Rightarrow \underbrace{\begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}}_{\mathbf{A}}$$

اگر هر کدام از درس ها را به صورت یک بردار در نظر بگیریم. ماتریس کواریانس را به صورت برداری زیر حساب میکنیم:

$$\mathbf{a} = \underbrace{\begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}}_{\text{ماتریس اصلی}} - \underbrace{\begin{bmatrix} 66 & 60 & 60 \\ 66 & 60 & 60 \\ 66 & 60 & 60 \\ 66 & 60 & 60 \\ 66 & 60 & 60 \end{bmatrix}}_{\text{ماتریس میانگین ها}} = \begin{bmatrix} 24 & 0 & 30 \\ 24 & 30 & -30 \\ -6 & 0 & 0 \\ -6 & 0 & 30 \\ -36 & -30 & -30 \end{bmatrix}$$

در مرحله اول هر کدام از نمونه ها را از میانگین کم کردیم. و در مرحله بعد حاصلضرب ها را محاسبه کردیم.نتیجه به صورت زیر می شود:

$$\mathbf{COV} = \mathbf{a}'\,\mathbf{a} / n\text{-}1 = \begin{bmatrix} 2520/4 & 1800/4 & 900/4 \\ 1800/4 & 1800/4 & 0/4 \\ 900/4 & 0/4 & 3600/4 \end{bmatrix} = \begin{bmatrix} 630 & 450 & 225 \\ 450 & 450 & 0 \\ 225 & 0 & 900 \end{bmatrix}$$

اندازه این ماتریس 3 در 3 است و به معنای ارتباط درس ها با هم است.

3- واریانس

واریانس به صورت «مقدار متوسط مربع اختلاف مقادیر از میانگین» تعریف شده است.

برای محاسبه واریانس، باید گام‌های زیر را دنبال کنید:

- ابتدا میانگین را پیدا کنید (میانگین ساده اعداد).
- سپس برای هر عدد، مقدار میانگین را از آن تفریق کرده و سپس نتیجه را به توان دو برسانید (مربع اختلاف).
- و در نهایت میانگین مربع اختلافات به دست آمده را محاسبه کنید.

مفهوم آماری واریانس (که عددی نامنفی‌ست) به ما می‌گوید که مقادیر یک متغیر تصادفی چگونه حول میانگین آن توزیع شده‌اند

$$\frac{f_1 \times (x_1 - \bar{x})^2 + f_2 \times (x_2 - \bar{x})^2 + \cdots + f_n \times (x_n - \bar{x})^2}{f_1 + f_2 + \cdots + f_n}$$

همینطور در جدول فراوانی دیدیم

$$\sigma^2 = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{n} f_i}$$

Variance and Standard Deviation Formula

cuemath
THE MATH EXPERT

	Population	Sample
Variance	$\sigma^2 = \dfrac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$	$S^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}$
Standard Deviation	$\sigma = \sqrt{\dfrac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}$	$S = \sqrt{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}}$

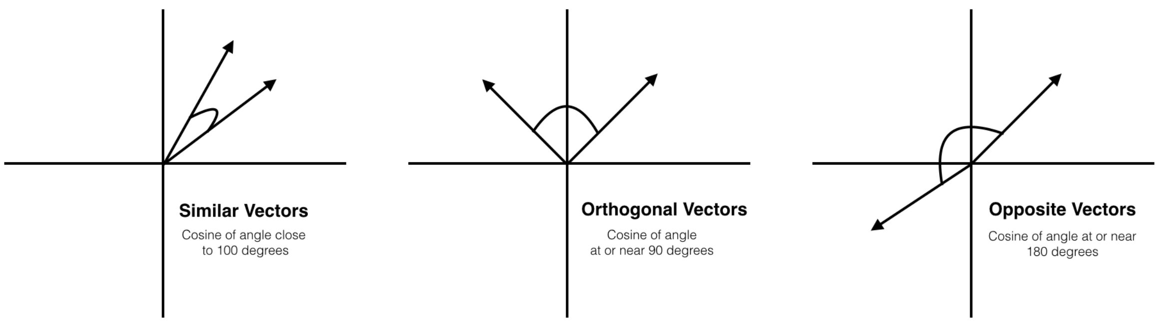
استاندارد دیویژن ، همانطور که در تصویر بالا مشخص است از جذرِ واریانس به دست می آید.

4- کوواریانس ، توضیح داده شد .

5 - شباهت کسینوسی :

معیارهای شباهت، معیارهایی مانند معیارهای فاصله هستند که میزان دور و یا نزدیک بودن دو موجودیت را مشخص می‌کنند. بدیهی است که معیار شباهت با معیارهای فاصله رابطه عکس دارند و به عبارتی هر چه میزان شباهت بیشتر باشد می‌توان نتیجه گرفت فاصله‌ی دو شیء کمتر است.

شباهت کسینوسی یک معیار شباهت است که پایه آن محاسبه‌ی میزان کسینوس زاویه‌ی بین دو بردار است.



در صورت انطباق دو بردار (در این معیار نشانه شباهت کامل است) که زاویه‌ی بین دو بردار صفر می‌باشد مقدار آن برابر ۱ خواهد شد و در کمترین میزان شباهت دو بردار یعنی اگر زاویه بین دو بردار ۱۸۰ درجه باشد نتیجه این معیار ۱- خواهد شد.

$$similarity(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}}$$

این معیار یکی از پرکاربردترین معیارها در پردازش متن است. معیار شباهت کسینوسی برای استفاده در زمانی که داده‌های مربوط به ویژگی‌های ما تنک یا sparse هستن مناسب است.

6- تابع چگالی ، توزیع نرمال :

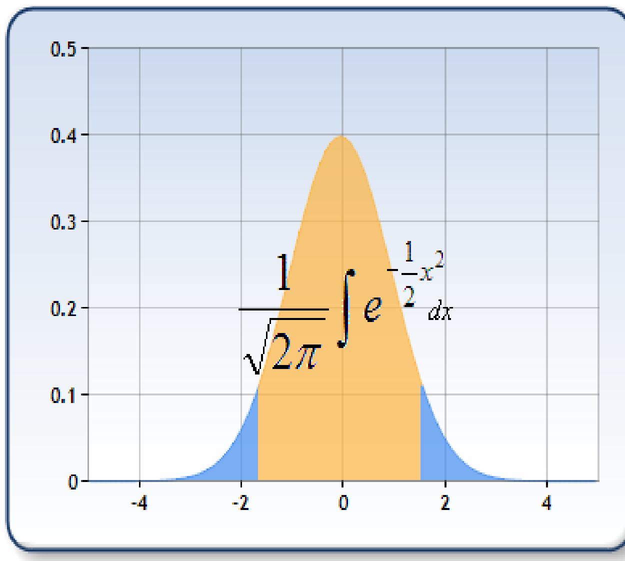
تابع چگالی احتمال توزیع نرمال با پارامترهای μ و σ^2 به صورت زیر است:

$$f(x;\,\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/(2\sigma^2)}, \qquad x \in \mathbb{R}.$$

Normal Distribution Formula

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$\mu =$ mean of x
$\sigma =$ standard deviation of x
$\pi \approx$ 3.14159 ...
$e \approx$ 2.71828 ...



شکل رسم شده پس از محاسبه توزیع نرمال باید مشابه شکل بالا باشد

توزیع نرمال یک تابع احتمال است که در آمار استفاده می شود و به نحوه توزیع مقادیر داده می گوید.
این تابع به دلیل مزایایی که در سناریوهای واقعی دارد، مهمترین تابع توزیع احتمال مورد استفاده در آمار است.
تابع چگالی احتمال برای توزیع نرمال بر حسب امید ریاضی و واریانس تعریف میشود.و تابع آن به صورت زیر است:

$$f(x)\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

اگر در این فرمول $\mu = 0, \sigma = 1$ باشد در این صورت به آن تابع **توزیع نرمال استاندارد** گویند. در این حالت تابع توزیع به صورت زیر خواهد بود:

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$$

فایل دیگری حاوی کد ها و خروجی و توضیح کد ها ضمیمه شده است. فرمت آن فایل به شکل html است با کلیک بر روی آن ، آن را در یک مرورگر باز نمایید. فایل مورد صحبت به کمک ژوپیتر نوت بوک ایجاد شده است.