

تشخیص آنومالی

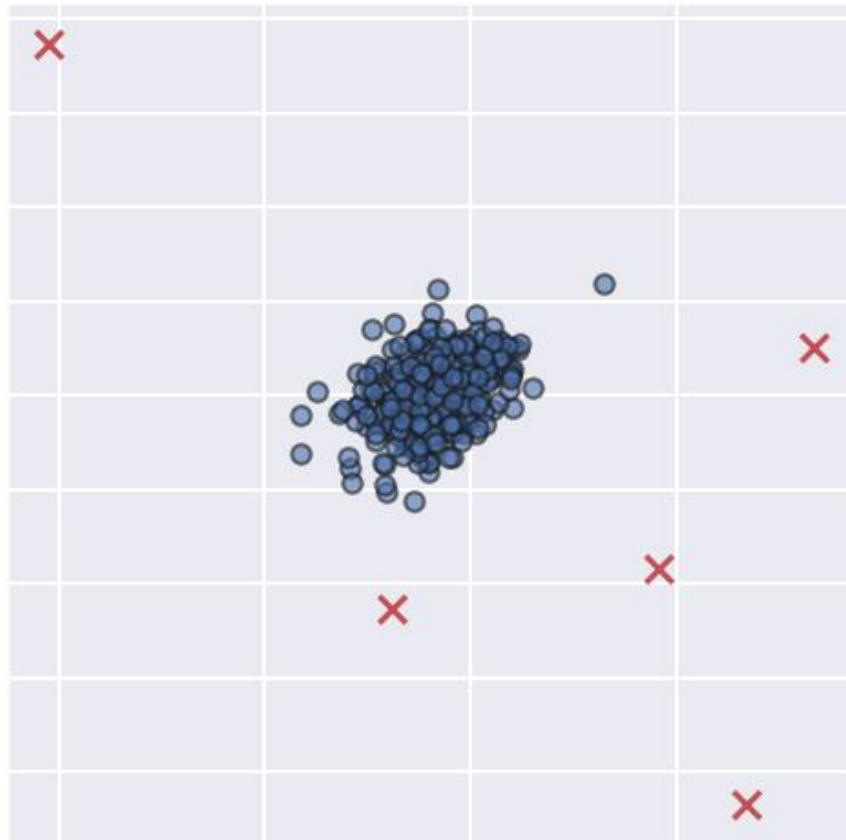
سید ناصر رضوی www.snrazavi.ir

۱۳۹۷

تشخیص آنومالی [تشخیص داده‌های پرت]

۲

□ تشخیص آنومالی. تشخیص مشاهداتی که با اغلب مشاهدات انجام شده به شدت تفاوت دارند.



□ تشخیص کلاهبرداری.

■ تشخیص تراکنش‌های بسیار **نامحتمل** توسط شخص دارای کارت اعتباری

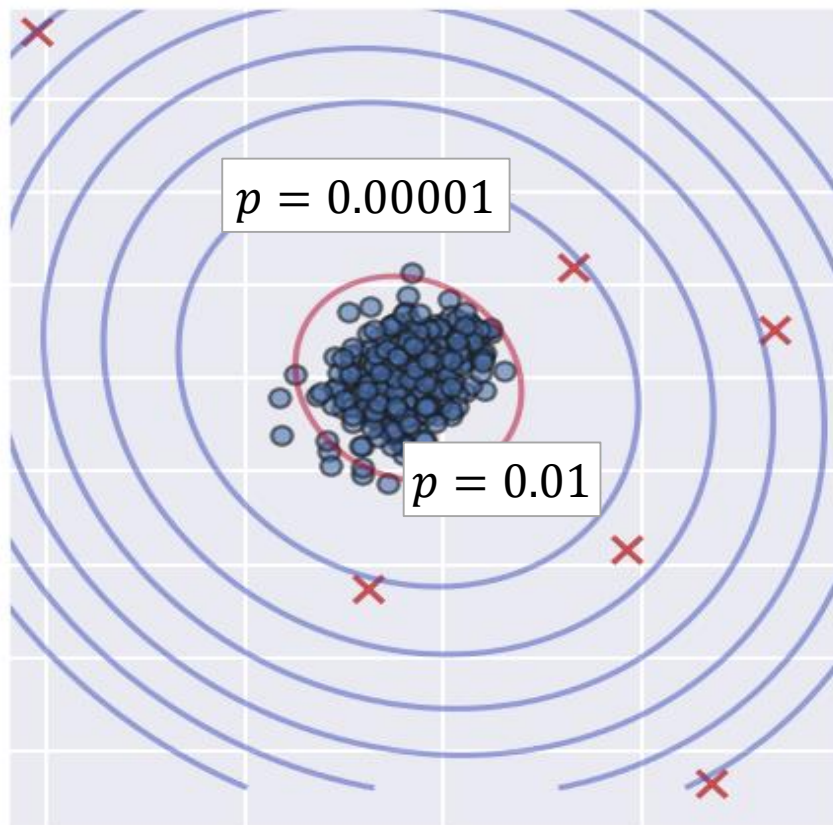
□ امنیت شبکه.

■ تشخیص فعالیت‌هایی که با **احتمال** بسیار کمی به وسیله یک کاربر قانونی انجام شده‌اند.

تشخیص آنومالی [تشخیص داده‌های پرت]

۳

□ تشخیص آنومالی. تشخیص مشاهداتی که با اغلب مشاهدات انجام شده به شدت تفاوت دارند.



□ رویکرد احتمالاتی برای تشخیص آنومالی.

□ ایجاد یک مدل احتمالاتی از داده‌ها

[بیانگر احتمال مشاهده هر رویداد ممکن]

□ مشخص کردن مشاهداتی که احتمال وقوع آنها بسیار کم است.

$$p(x) < \epsilon$$

توزیع گوسی (نرمال)

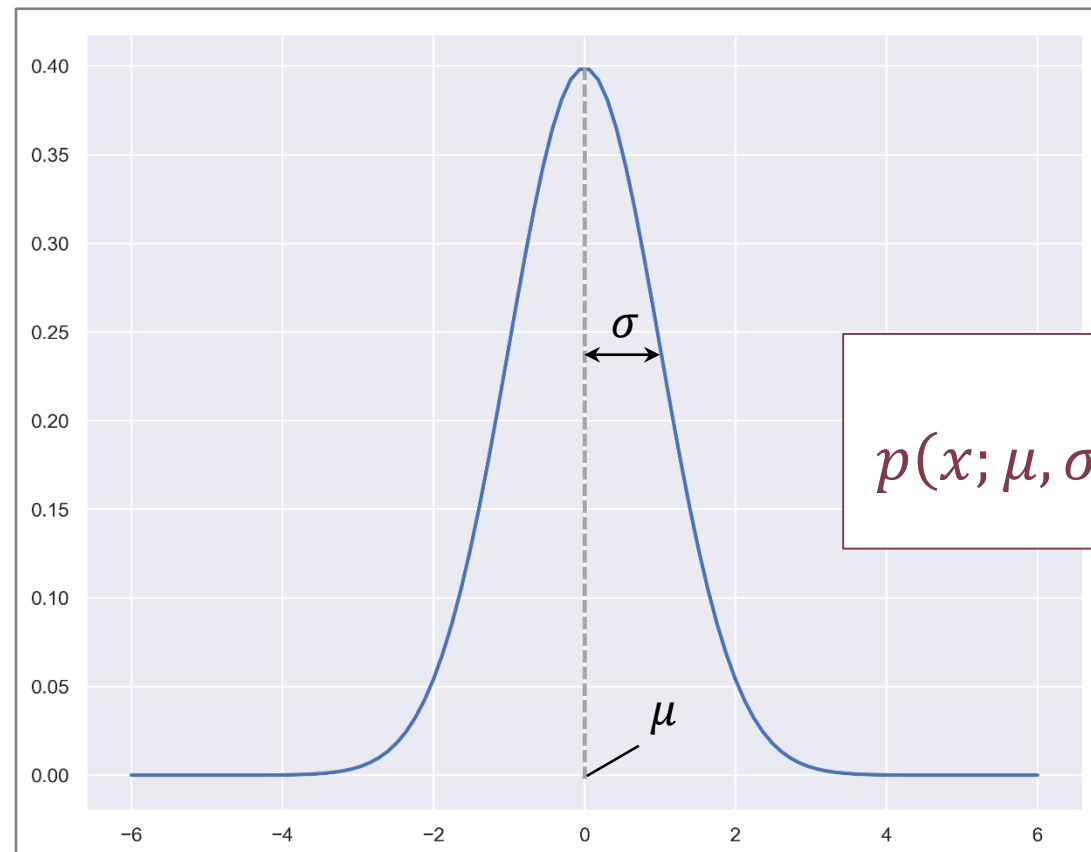


توزیع گوسی

۵

□ توزیع گوسی. فرض کنید x دارای توزیع گوسی با میانگین μ و واریانس σ^2 باشد.

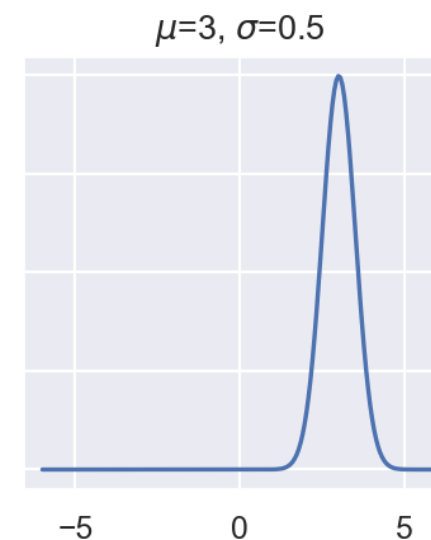
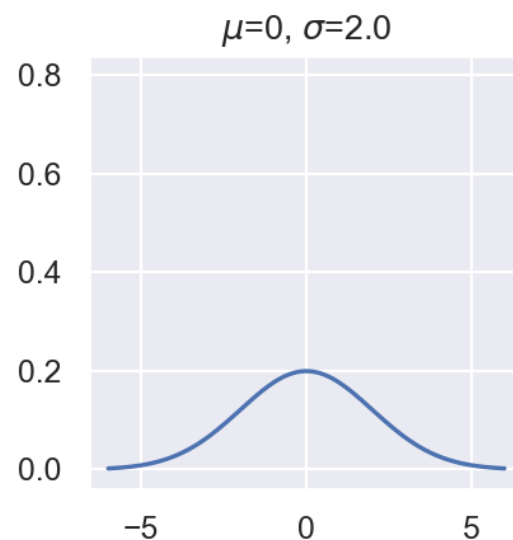
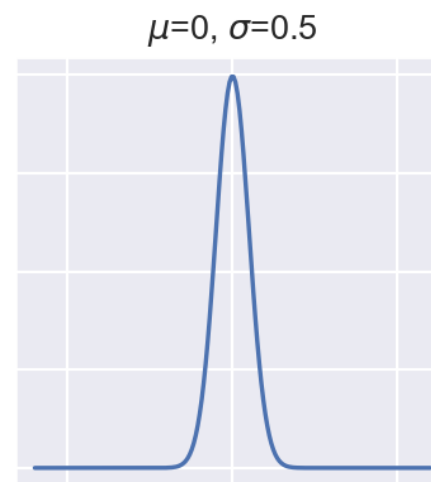
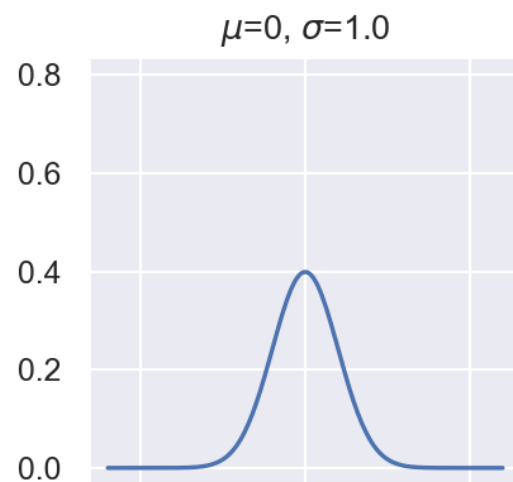
$$x \sim \mathcal{N}(\mu, \sigma^2)$$



$$p(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

توزیع گوسی تک متغیره

۶



تخمین پارامتر

۷

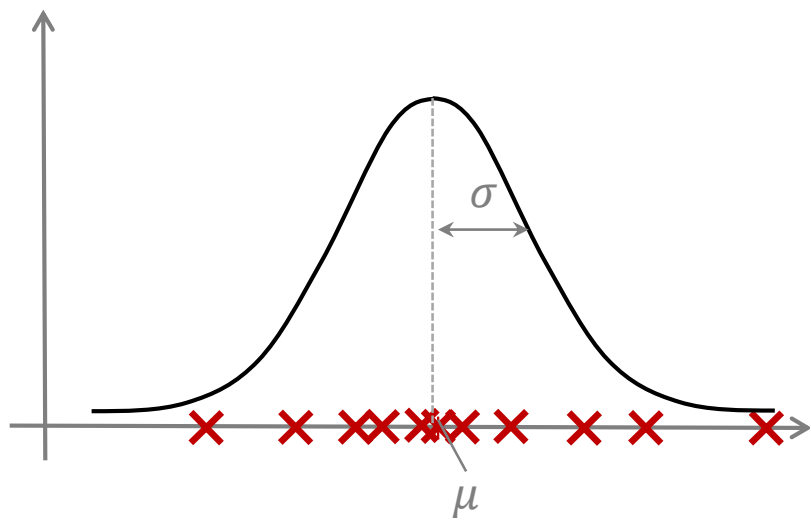
□ مجموعه داده.

$$\{x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(m)}\}$$

□ هدف. تخمین مقادیر μ و σ

$$\mu = \frac{1}{m} \sum_{i=1}^m x^{(i)}$$

$$\sigma^2 = \frac{1}{m} \sum_{i=1}^m (x^{(i)} - \mu)^2$$



الڳوريتم تشخيص آنومالي



الگوریتم تخمین توزیع

۹

□ مجموعه آموزشی.

$$\{x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(m)}\}, \quad x^{(i)} \in \mathbb{R}^n$$

□ فرضیات.

□ ویژگی‌ها از توزیع نرمال پیروی می‌کنند.
 $x_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$

□ بین ویژگی‌ها همبستگی وجود ندارد. [ماتریس کوواریانس قطری است]

$$\begin{aligned} p(\mathbf{x}) &= p(x_1; \mu_1, \sigma_1^2) p(x_2; \mu_2, \sigma_2^2) p(x_3; \mu_3, \sigma_3^2) \cdots p(x_n; \mu_n, \sigma_n^2) \\ &= \prod_{j=1}^n p(x_j; \mu_j, \sigma_j^2) \end{aligned}$$

الگوریتم تشخیص آنومالی

۱۰

□ تعیین ویژگی‌هایی که می‌توانند در تشخیص آنومالی مفید باشند.

□ تخمین پارامترها (به ازای $1 \leq j \leq n$)

$$\mu_j = \frac{1}{m} \sum_{i=1}^m x_j^{(i)} \quad \sigma_j^2 = \frac{1}{m} \sum_{i=1}^m \left(x_j^{(i)} - \mu_j \right)^2$$

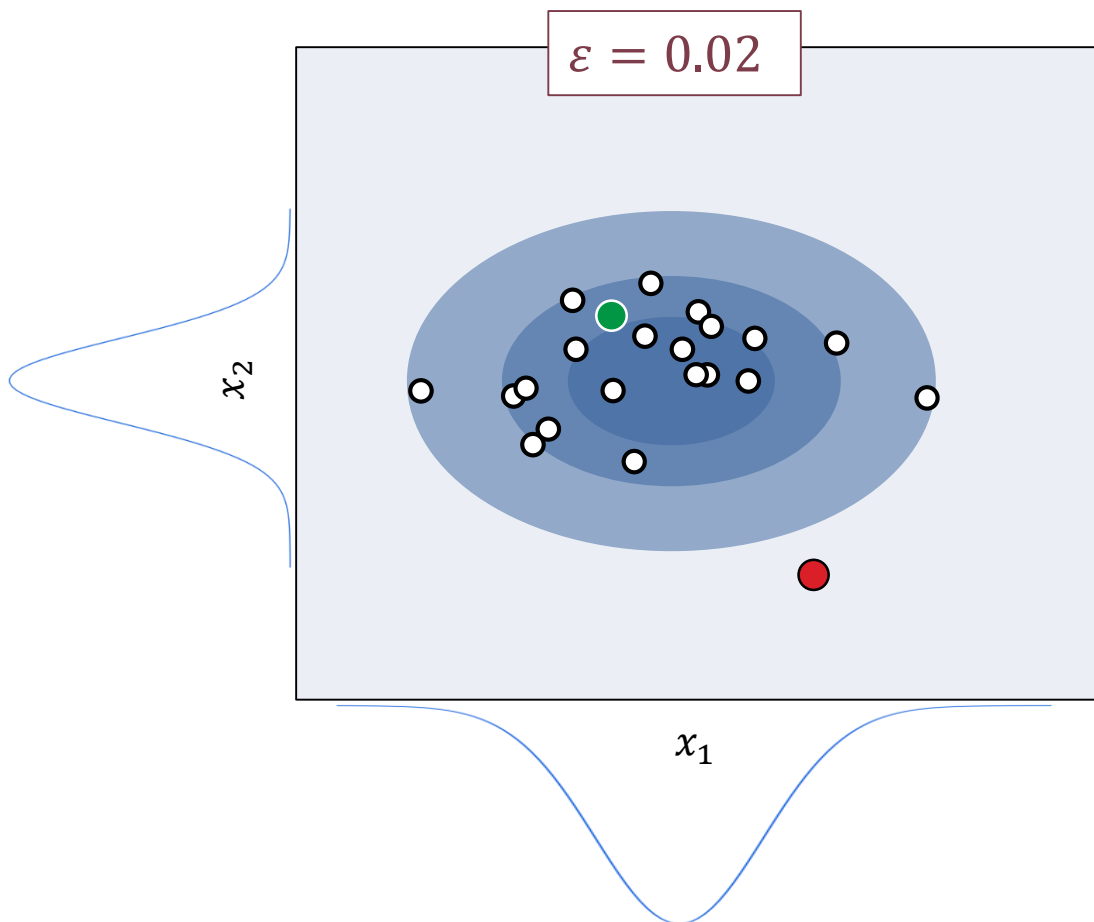
□ محاسبه $p(x)$ برای داده جدید x

$$p(x) = \prod_{j=1}^n p(x_j; \mu_j, \sigma_j^2) = \prod_{j=1}^n \frac{1}{\sigma_j \sqrt{2\pi}} \exp \left(-\frac{(x_j - \mu_j)^2}{2\sigma_j^2} \right)$$

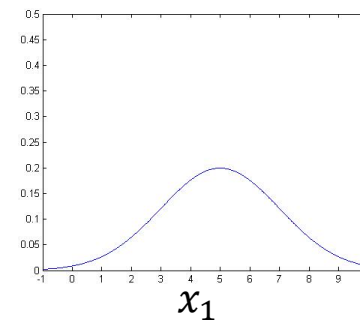
□ تولید خروجی «بله» به شرطی که $p(x) < \epsilon$

مثال

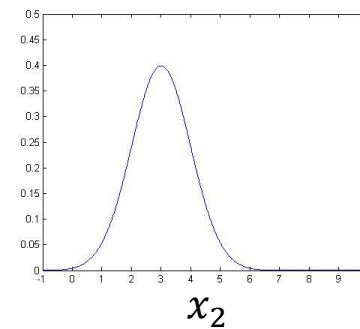
۱۱



$$\mu_1 = 5, \sigma_1 = 2$$



$$\mu_2 = 3, \sigma_2 = 1$$



$$p(x_{test}^{(1)}) = 0.0426$$

$$p(x_{test}^{(2)}) = 0.0021$$

توسعه و ارزیابی سیستم‌های تشخیص آنومالی

ارزیابی‌های عددی

۱۳

□ اهمیت.

□ در طول فرآیند توسعه سیستم‌های یادگیری، اگر روشی به منظور ارزیابی سیستم در اختیار داشته باشیم، آنگاه بسیاری از تصمیم‌گیری‌ها (همانند انتخاب ویژگی‌ها و غیره) بسیار ساده‌تر خواهد شد.

□ فرض کنید تعدادی داده برچسب‌گذاری شده داریم به طوری که به ازای هر داده، عادی بودن ($y = 0$) و یا غیرعادی بودن ($y = 1$) آن مشخص شده است.

□ مجموعه آموزشی. [شامل داده‌های عادی]

$$\{x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(m)}\}$$

□ مجموعه اعتبارسنجی.

$$\left\{ \left(x_{cv}^{(1)}, y_{cv}^{(1)} \right), \left(x_{cv}^{(2)}, y_{cv}^{(2)} \right), \dots, \left(x_{cv}^{(m_{cv})}, y_{cv}^{(m_{cv})} \right) \right\}$$

□ مجموعه آزمایشی.

$$\left\{ \left(x_{test}^{(1)}, y_{test}^{(1)} \right), \left(x_{test}^{(2)}, y_{test}^{(2)} \right), \dots, \left(x_{test}^{(m_{test})}, y_{test}^{(m_{test})} \right) \right\}$$

□ مجموعه داده. اطلاعات مربوط به عملکرد موتورها

□ ۱۰۰۰۰ موتور سالم

□ ۲۰ موتور معیوب

□ تقسیم‌بندی داده‌ها.

□ مجموعه آموزشی:

□ مجموعه اعتبارسنجی:

□ مجموعه آزمایشی.

۶۰۰۰ موتور سالم دسته‌بندی تک‌دسته‌ای

۲۰۰۰ موتور سالم و ۱۰ موتور معیوب

۲۰۰۰ موتور سالم و ۱۰ موتور معیوب

ارزیابی الگوریتم

۱۵

□ آموزش. توسعه مدل $p(x)$ با توجه به مجموعه آموزشی

□ پیش‌بینی. برای نمونه‌های موجود در مجموعه اعتبارسنجی یا آموزشی

$$y = \begin{cases} 1, & p(x) < \varepsilon \\ 0, & p(x) \geq \varepsilon \end{cases}$$

□ معیارهای ارزیابی ممکن.

□ مثبت درست، مثبت غلط، منفی درست، منفی غلط

□ نرخ دقت و نرخ یادآوری

□ امتیاز F_1

□ توجه. می‌توان از مجموعه اعتبارسنجی به منظور انتخاب یک مقدار مناسب برای ε استفاده نمود.

معیارهای ارزیابی

۱۶

□ معیارهای ارزیابی. برای داده‌های نامتوازن

پیش‌بینی	واقعی	
	$y = 1$	$y = 0$
	$y = 1$	$y = 0$
	TP	FP
	FN	TN

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \frac{P \cdot R}{P + R}$$

تشخیص آنومالی یا یادگیری نظارت شده؟

تشخیص آنومالی یا یادگیری نظارت شده؟

۱۸

تشخیص آنومالی

□ تعداد نمونه‌ها.

□ نسبت تعداد نمونه‌های مثبت به منفی بسیار کم

□ «انواع» بسیار متفاوت از آنومالی‌ها.

□ برای هر الگوریتمی، یاد گرفتن آنومالی‌ها از روی تعداد کم نمونه‌های مثبت بسیار دشوار است.

□ آنومالی‌های جدید ممکن است هیچ شباهتی به آنومالی‌هایی که قبلاً دیده شده‌اند، نداشته باشند.

یادگیری نظارت شده

□ تعداد نمونه‌ها.

□ تعداد نمونه‌های مثبت و منفی زیاد

□ نمونه‌های مثبت.

□ تعداد نمونه‌های مثبت برای این که الگوریتم درک درستی از نمونه‌های مثبت پیدا کند، کفایت می‌کند.

□ نمونه‌های مثبت جدید به احتمال زیاد شبیه به نمونه‌های مثبتی هستند که الگوریتم قبلاً در طی فرآیند آموزش با آنها مواجه شده است.

تشخیص آنومالی یا یادگیری نظارت شده؟

۱۹

تشخیص آنومالی

□ تشخیص کلاهبرداری.

□ ساخت و تولید (ساختن موتور هواپیما).

□ نظارت بر ماشین‌ها در مراکز داده‌ای.

□ ...

یادگیری نظارت شده

□ تشخیص هرزنامه.

□ پیش‌بینی وضعیت آب و هوا.

□ تشخیص غده‌های سرطانی بدخیم.

□ ...

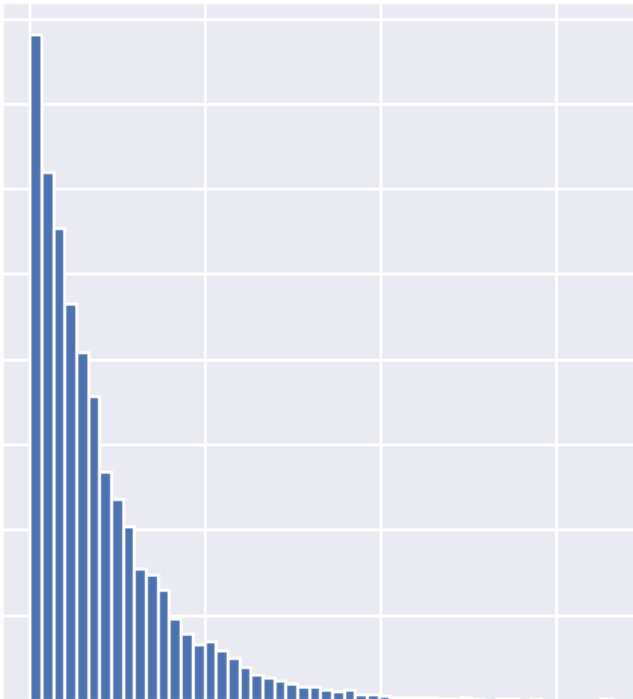
انتخاب ویژگی‌ها



تبدیل ویژگی با توزیع غیرنرمال به ویژگی با توزیع نرمال

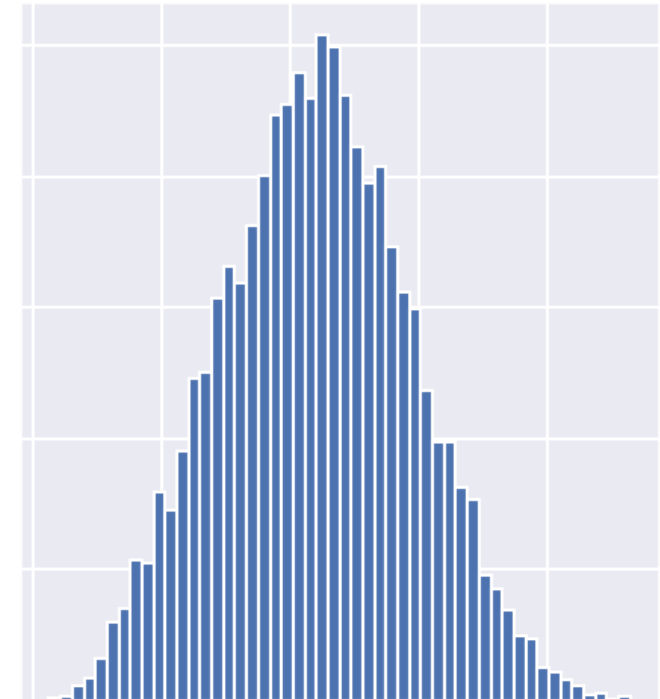
۲۱

Original Data (Gamma Distribution)



$$x^{0.3}$$

Transformed Data (Normal Distribution)



```
x = np.random.gamma(1, 2, (10000, 1))  
plt.hist(x, 50)
```

```
plt.hist(x ** 0.3, 50)
```

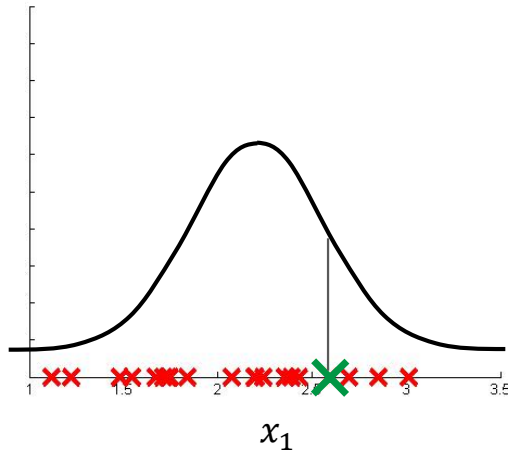
تحلیل خطا برای کمک به تشخیص آنومالی

۲۲

□ هدف. می‌خواهیم مقدار $p(x)$:

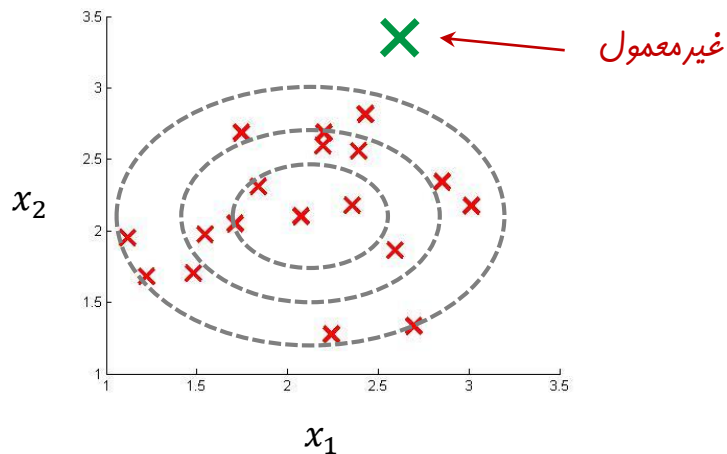
□ برای داده‌های عادی بزرگ باشد.

□ برای داده‌های غیرعادی کوچک باشد.



□ یک مشکل متداول.

□ $p(x)$ برای داده‌های عادی و غیرعادی تفاوت چندانی ندارد.



نظارت بر کامپیوترها در مراکز داده‌ای

۲۳

□ انتخاب ویژگی‌ها. انتخاب ویژگی‌هایی که در صورت وجود آنومالی مقدارشان **بسیار کوچک** یا **بسیار بزرگ** باشد.

□ میزان مصرف حافظه

□ تعداد دستیابی‌ها به دیسک در ثانیه

□ میزان بار پردازنده

□ ترافیک شبکه

□ افزودن ویژگی‌های جدید برای تشخیص شرایط غیرعادی.

□ نسبت بار پردازنده به ترافیک شبکه

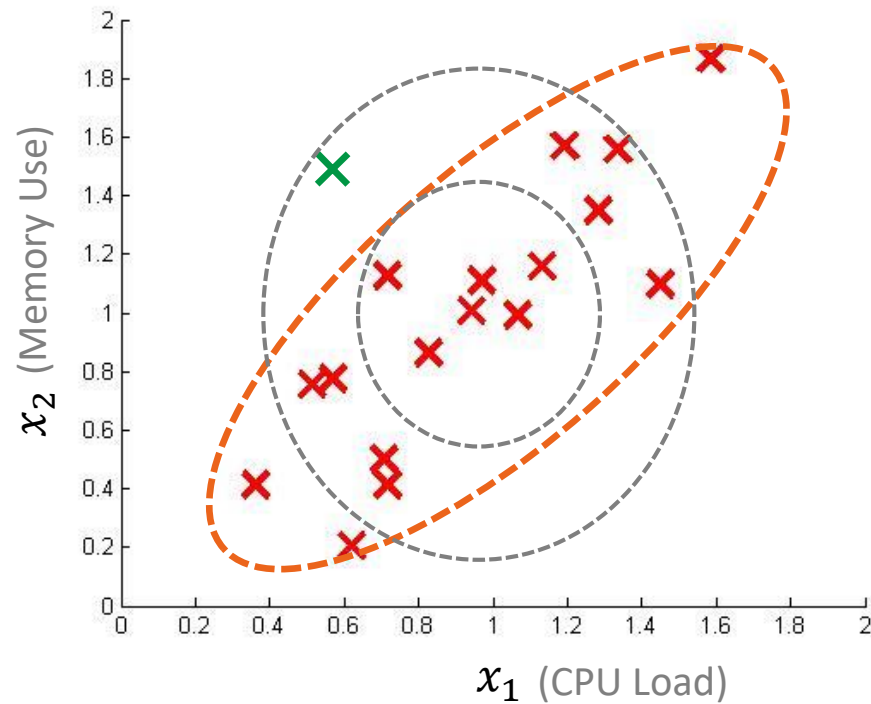
[مثلا اگر پردازنده در یک حلقه بی‌نهایت گیر کرده باشد، مقدار این ویژگی بسیار بزرگ خواهد بود]

توزیع گوسی چندمتغیره

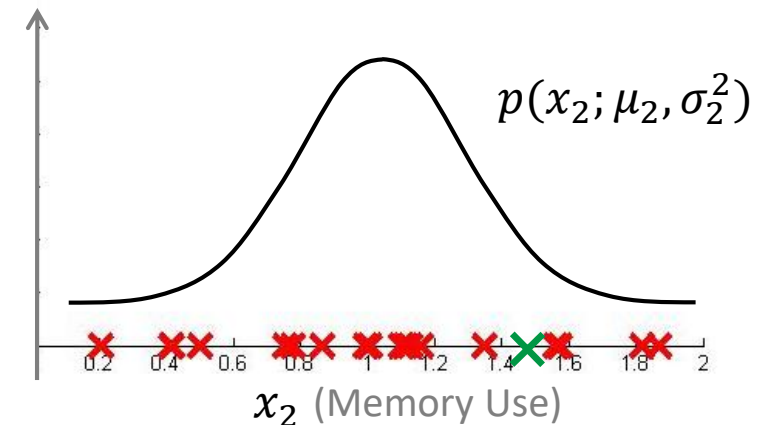
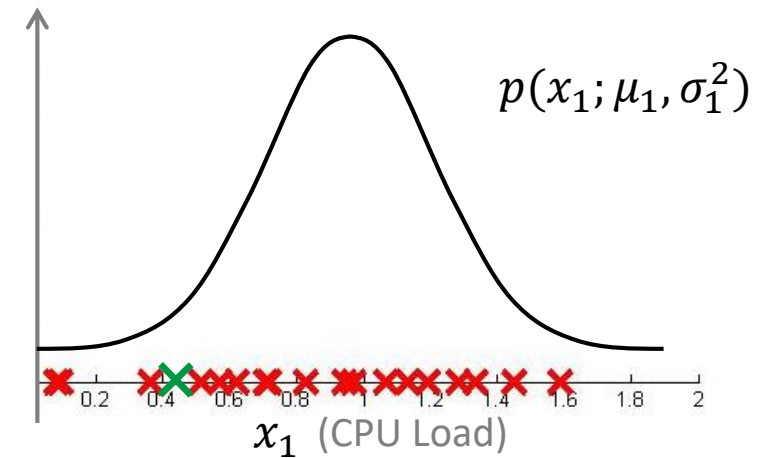
مثال مقدماتی

۲۵

تابع گوسی دومتغیره



با افزایش بار پردازنده، مصرف حافظه به طور معمول بیشتر می شود. (وجود همبستگی میان ویژگی ها)



تابع گوسی چند متغیره

۲۶

□ تابع گوسی چند متغیره.

$$p(x; \mu, \Sigma) = \frac{1}{|\Sigma|^{1/2} (2\pi)^{n/2}} \exp \left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right)$$

□ پارامترها.

ماتریس کوواریانس

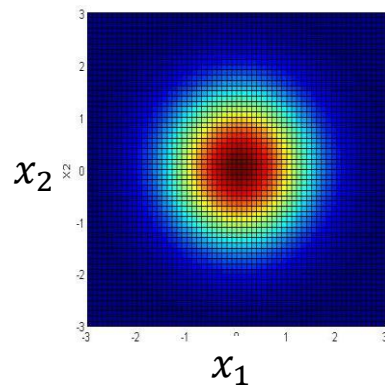
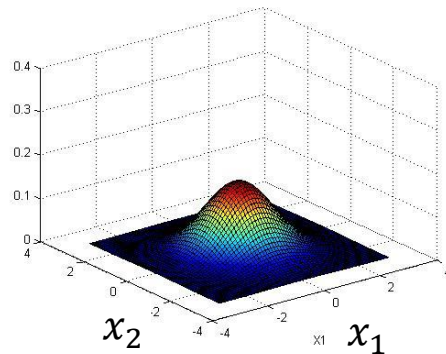
$$\Sigma \in \mathbb{R}^{n \times n}$$

$$\mu \in \mathbb{R}^n$$

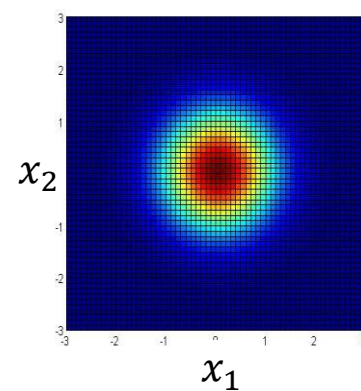
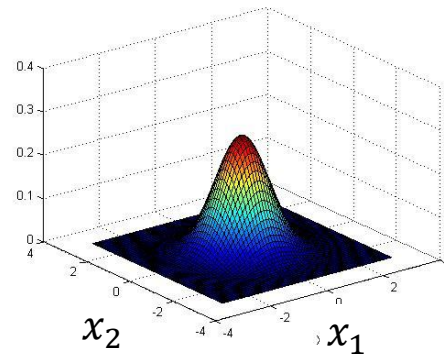
ماتریس کوواریانس قطری، واریانس ویژگی‌ها برابر

۲۷

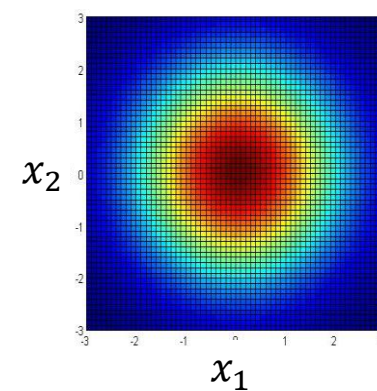
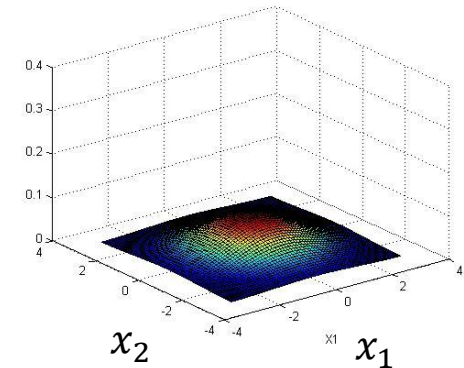
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 0.6 & 0 \\ 0 & 0.6 \end{bmatrix}$$



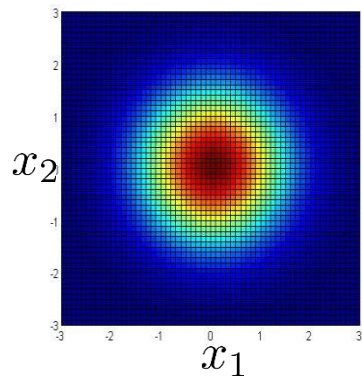
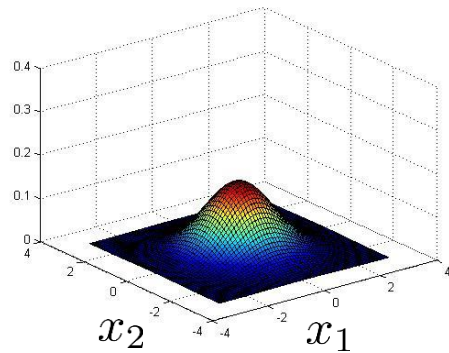
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$



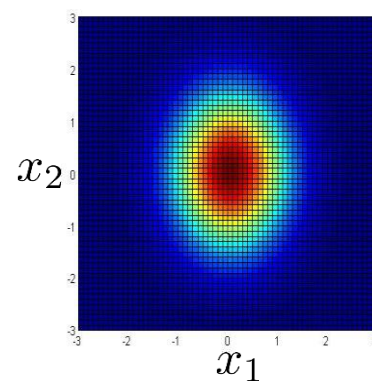
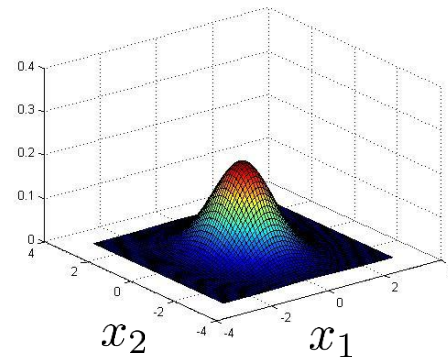
ماتریس کوواریانس قطری، واریانس ویژگی‌ها متفاوت

۲۸

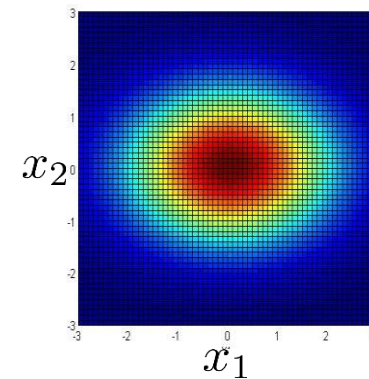
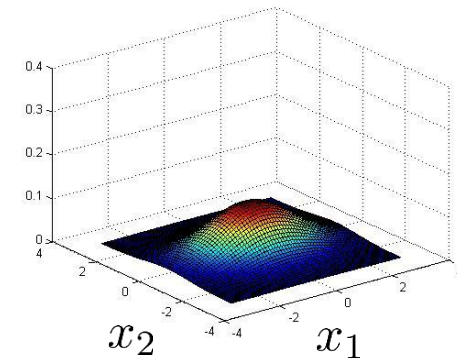
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 0.6 & 0 \\ 0 & 1 \end{bmatrix}$$



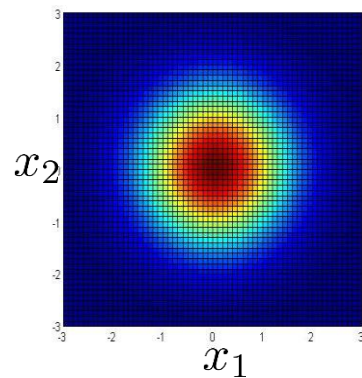
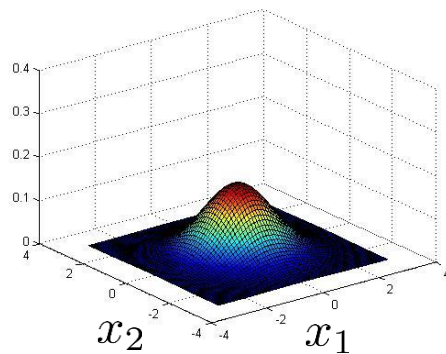
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$



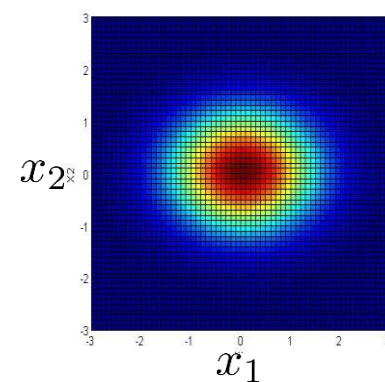
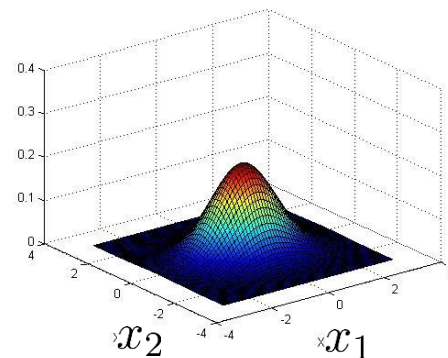
ماتریس کوواریانس قطری، واریانس ویژگی‌ها متفاوت

۲۹

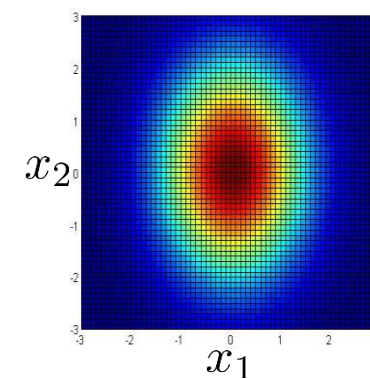
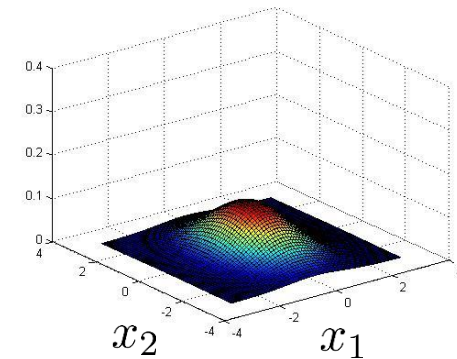
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 0.6 \end{bmatrix}$$



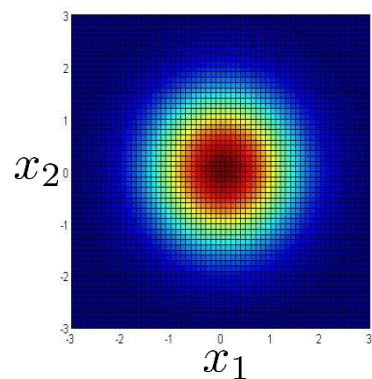
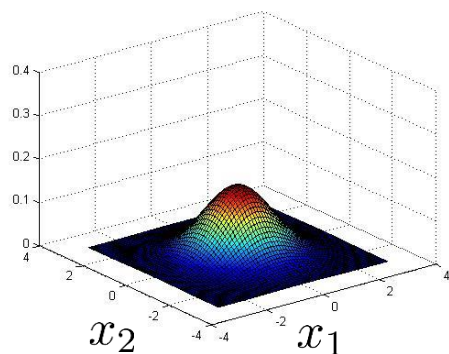
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$



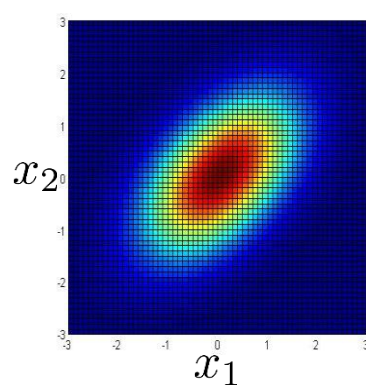
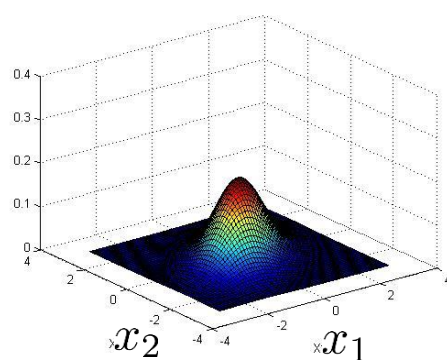
وجود همبستگی مثبت میان ویژگی‌ها

۳۰

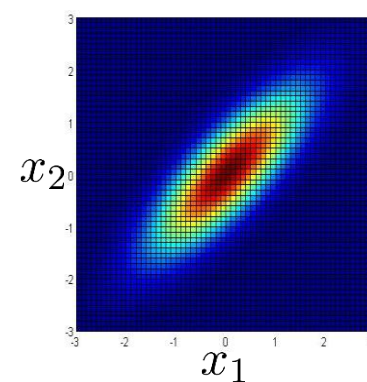
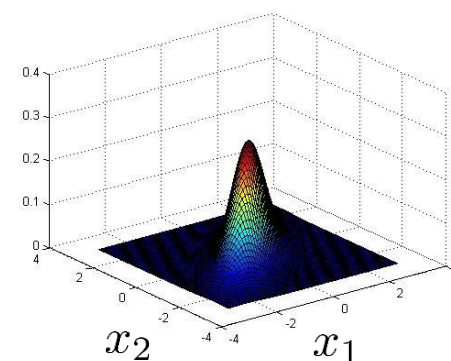
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$



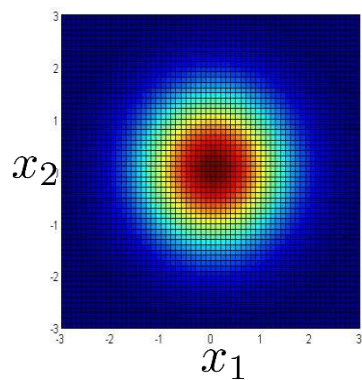
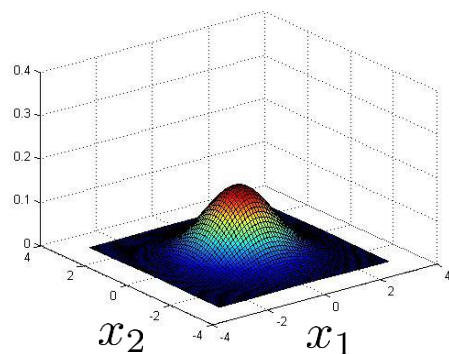
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$



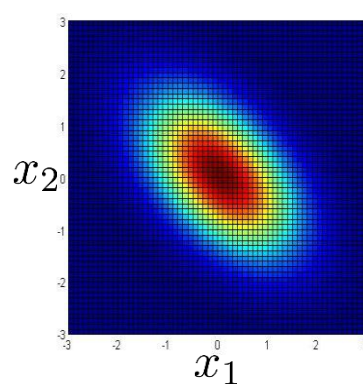
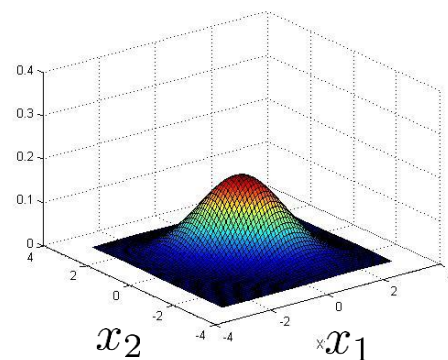
وجود همبستگی منفی ویژگی‌ها

۳۱

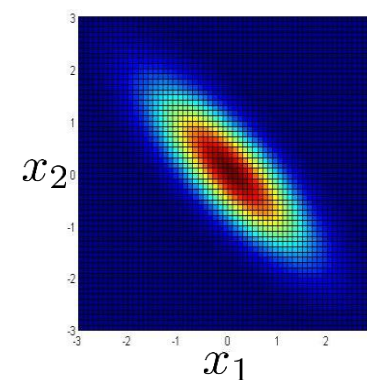
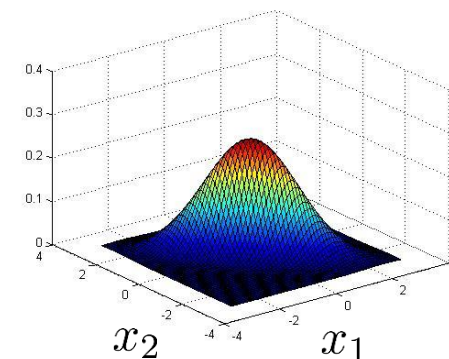
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$$



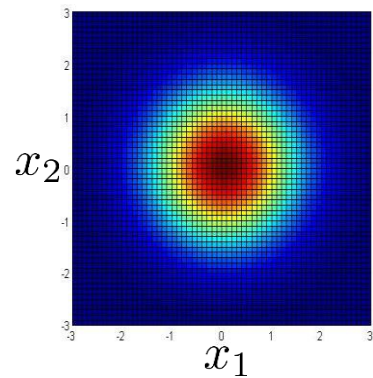
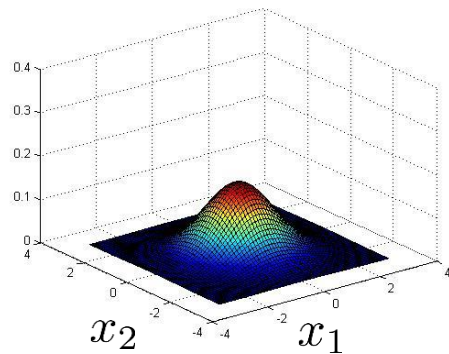
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}$$



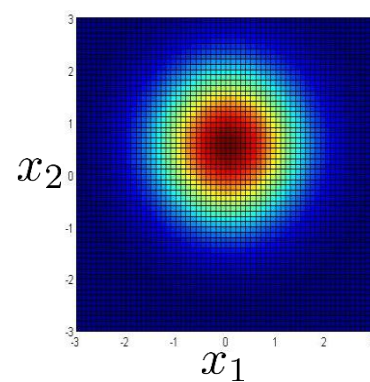
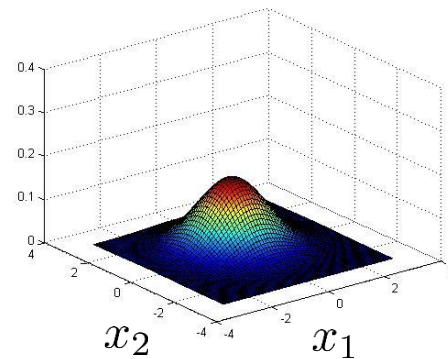
مرکز (میانگین) توزیع گاوسی

۳۲

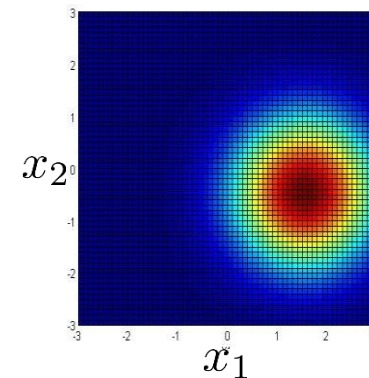
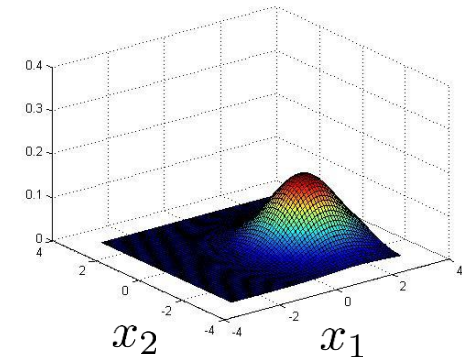
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\mu = \begin{bmatrix} 0 \\ 0.5 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\mu = \begin{bmatrix} 1.5 \\ -0.5 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



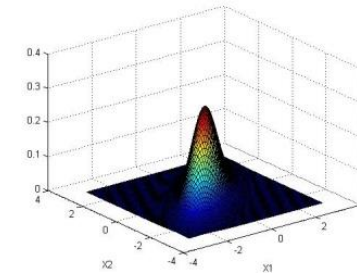
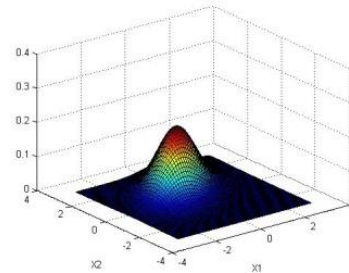
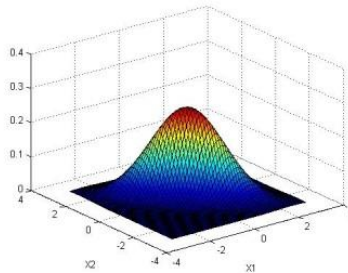
تشخیص آنومالی با تابع گوسی چند متغیره

توزیع گوسی چند متغیره

۳۴

□ تابع توزیع گوسی چند متغیره.

$$p(x; \mu, \Sigma) = \frac{1}{|\Sigma|^{1/2} (2\pi)^{n/2}} \exp \left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right)$$



□ تخمین پارامترها.

$$\mu = \frac{1}{m} \sum_{i=1}^m x^{(i)}$$

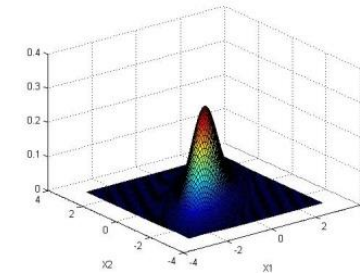
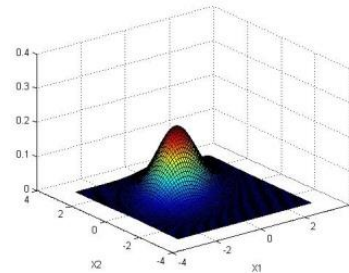
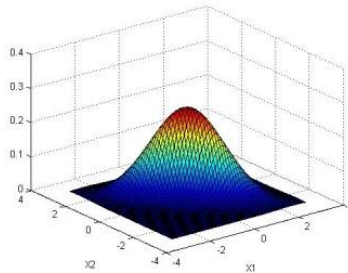
$$\Sigma = \frac{1}{m} \sum_{i=1}^m (x^{(i)} - \mu)(x^{(i)} - \mu)^T$$

توزیع گوسی چند متغیره

۳۵

□ تابع توزیع گوسی چند متغیره.

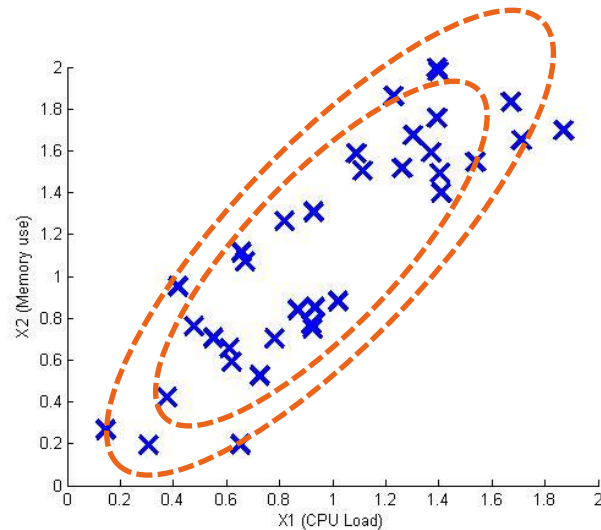
$$p(x; \mu, \Sigma) = \frac{1}{|\Sigma|^{1/2} (2\pi)^{n/2}} \exp \left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right)$$



□ تخمین پارامترها.

```
mu = np.mean(X, axis=0)
```

```
Sigma = np.cov(X.T)
```



□ تخمین پارامترهای مدل $p(x)$

$$\mu = \frac{1}{m} \sum_{i=1}^m x^{(i)}$$

$$\Sigma = \frac{1}{m} \sum_{i=1}^m (x^{(i)} - \mu)(x^{(i)} - \mu)^T$$

□ محاسبه مقدار $p(x)$ برای داده‌ی جدید x

$$p(x; \mu, \Sigma) = \frac{1}{|\Sigma|^{1/2} (2\pi)^{n/2}} \exp \left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right)$$

□ تولید خروجی «بله» اگر $p(x) < \varepsilon$

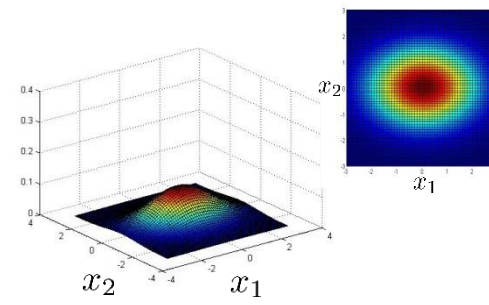
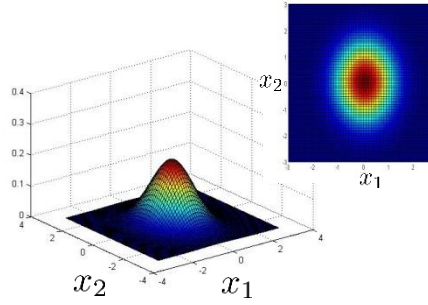
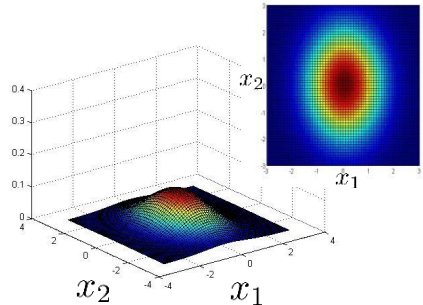
رابطه با مدل اولیه

۳۷

□ مدل اولیه.

$$p(\mathbf{x}) = p(x_1; \mu_1, \sigma_1^2) p(x_2; \mu_2, \sigma_2^2) p(x_3; \mu_3, \sigma_3^2) \cdots p(x_n; \mu_n, \sigma_n^2)$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$



□ رابطه با توزیع گوسی چند متغیره.

$$p(\mathbf{x}; \mu, \Sigma) = \frac{1}{|\Sigma|^{1/2} (2\pi)^{n/2}} \exp \left(-\frac{1}{2} (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \right)$$

مدل اولیه یا توزیع گوسی چند متغیره

۳۸

□ مدل اولیه.

- ایجاد ویژگی‌ها به صورت دستی انجام می‌شود. (x_1/x_2)
- هزینه‌های محاسباتی به نسبت پایین است.
- اگر تعداد نمونه‌های آموزشی کم باشد، باز هم به درستی عمل می‌کند. [تعداد پارامترها: $2n$]

□ توزیع گوسی چند متغیره.

- به طور خودکار همبستگی میان ویژگی‌ها را یاد می‌گیرد.
- هزینه‌های محاسباتی بالا است. [محاسبه واریانس کوواریانس]
- تعداد نمونه‌های آموزشی باید از تعداد ویژگی‌ها بیشتر باشد. [واریانس پذیری ماتریس Σ]