ECE 417 Lecture 8: Gaussians

Mark Hasegawa-Johnson

9/21/2021

Contents

- Gaussian pdf; Central limit theorem, Brownian motion
- White Noise
- Vector of i.i.d. Gaussians
- Vector of Gaussians that are independent but not identical
- Facts about linear algebra
- Vector of Gaussians that are neither independent nor identical

Review: Bayesian Classifier

A Bayesian classifier computes

$$h(x) = \text{argmax}\, p_{Y|X}(y|x) = \text{argmax}\, p_Y(y) p_{X|Y}(x|y)$$

- The prior, $p_Y(y)$ is just a lookup table, but…
- The likelihood, $p_{X|Y}(x|y)$, usually needs to be some kind of parameterized pdf. A Gaussian is often an excellent choice.

Gaussian (Normal) pdf

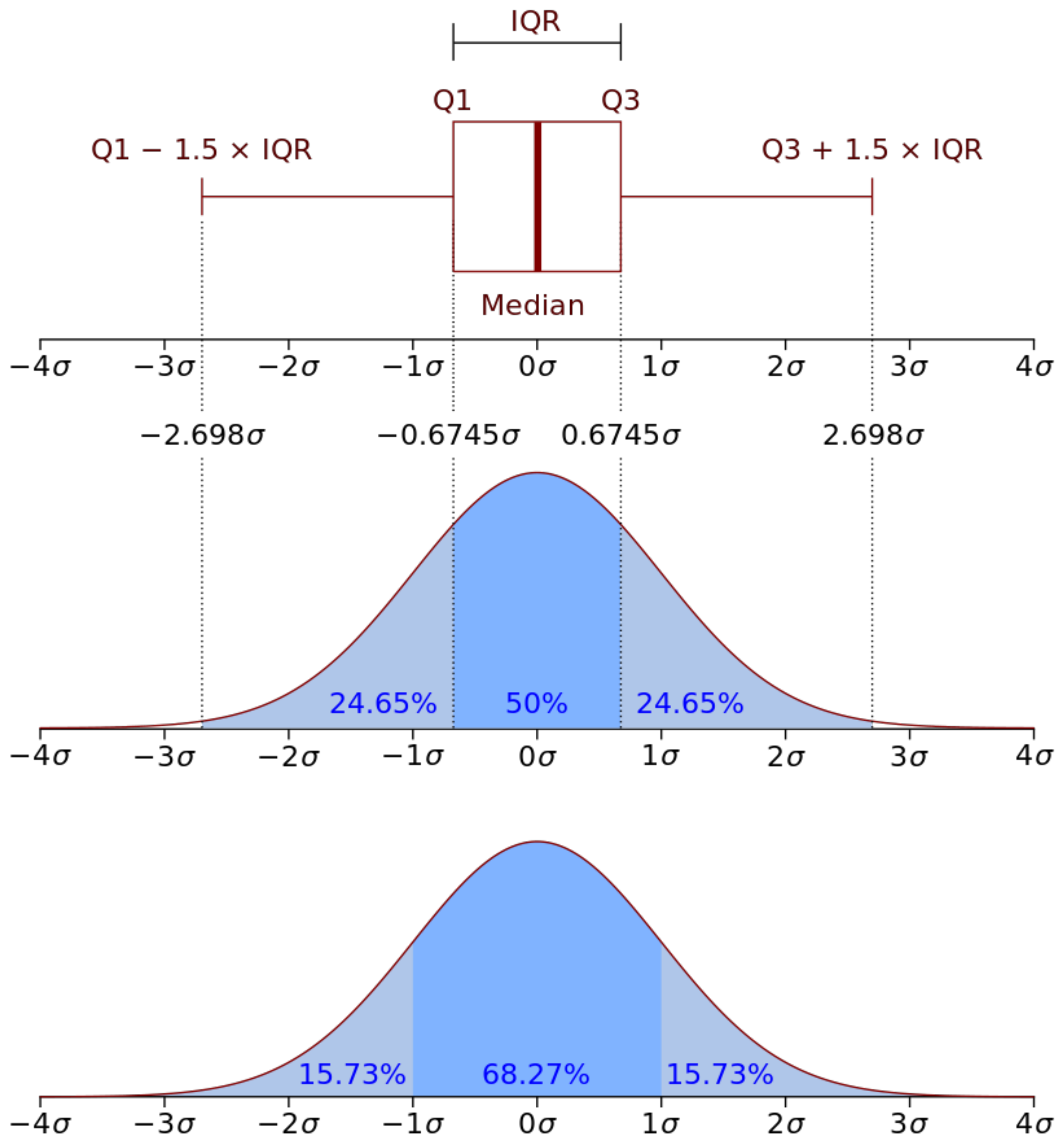
Gauss considered this problem: under what circumstances does it make sense to estimate the mean of a distribution, μ, by taking the average of the experimental values, $\mathrm{m} = \frac{1}{n}\sum_{i=1}^{n} x_i$?

He demonstrated that m is the maximum likelihood estimate of μ if (not only if!) X is distributed with the following probability density:

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

Gaussian pdf

Attribution: jhguch, https://commons.wikimedia.org/wiki/File:Boxplot_vs_PDF.svg



Unit Normal pdf

Suppose that X is normal with mean μ and standard deviation σ (variance σ^2):

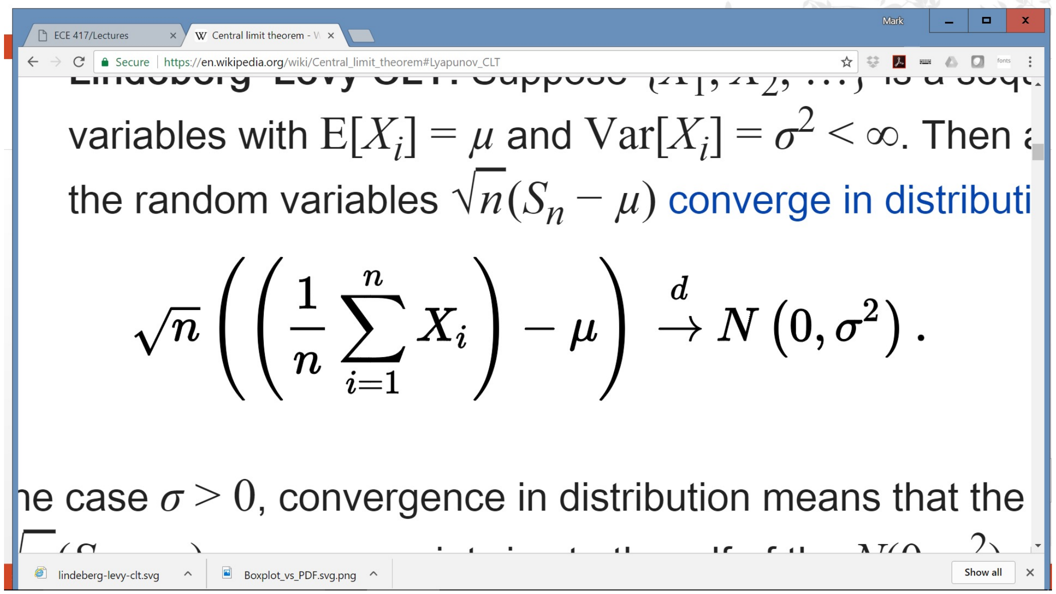
$$p_X(x) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Then $U = \left(\frac{X-\mu}{\sigma}\right)$ is normal with mean 0 and standard deviation 1:

$$p_U(u) = \mathcal{N}(u; 0,1\ ) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$$

Central Limit Theorem

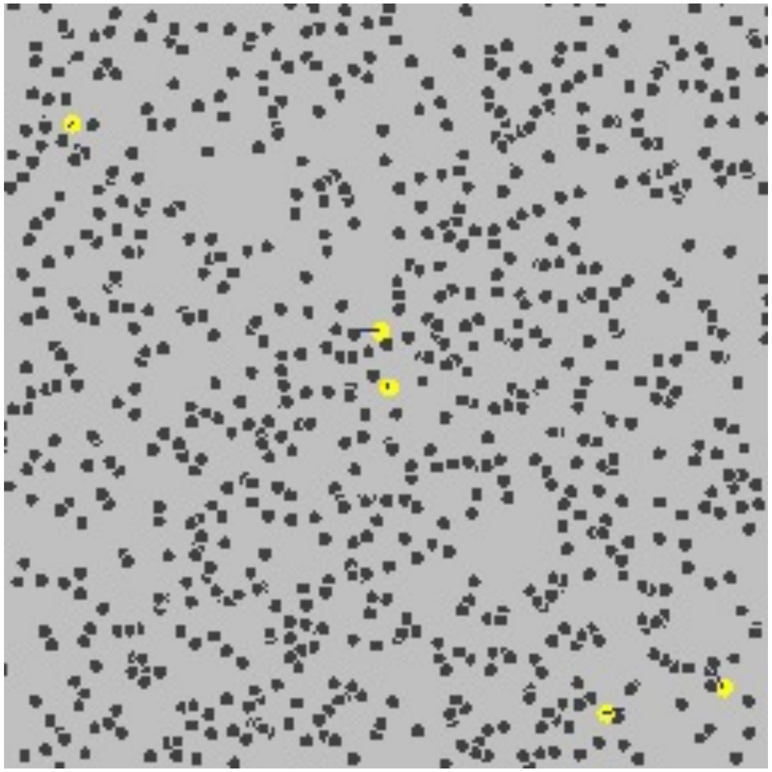
The Gaussian pdf is important because of the Central Limit Theorem. Suppose X_i are i.i.d. (independent and identically distributed), each having mean μ and variance σ^2. Then



Brownian motion

The Central Limit Theorem matters because Einstein showed that the movement of molecules, in a liquid or gas, is the sum of n i.i.d. molecular collisions.

In other words, the position after t seconds is Gaussian, with mean 0, and with a variance of Dt, where D is some constant.



Attribution: lookang,
https://commons.wikimedia.org/wiki/File:Brownianmotion5particles150frame.gif

Contents

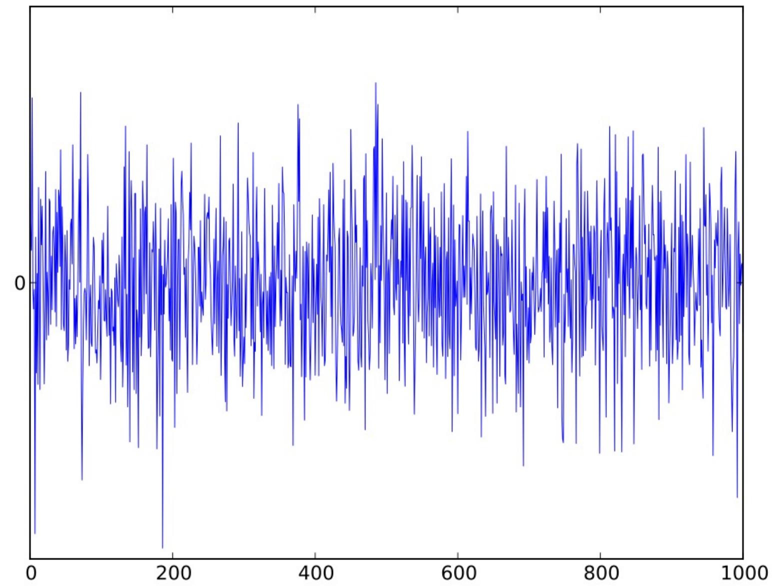
- Gaussian pdf; Central limit theorem, Brownian motion
- White Noise
- Vector of i.i.d. Gaussians
- Vector of Gaussians that are independent but not identical
- Facts about linear algebra
- Vector of Gaussians that are neither independent nor identical

Gaussian Noise

- Sound = air pressure fluctuations caused by velocity of air molecules
- Velocity of warm air molecules without any external sound source = Gaussian

Therefore:

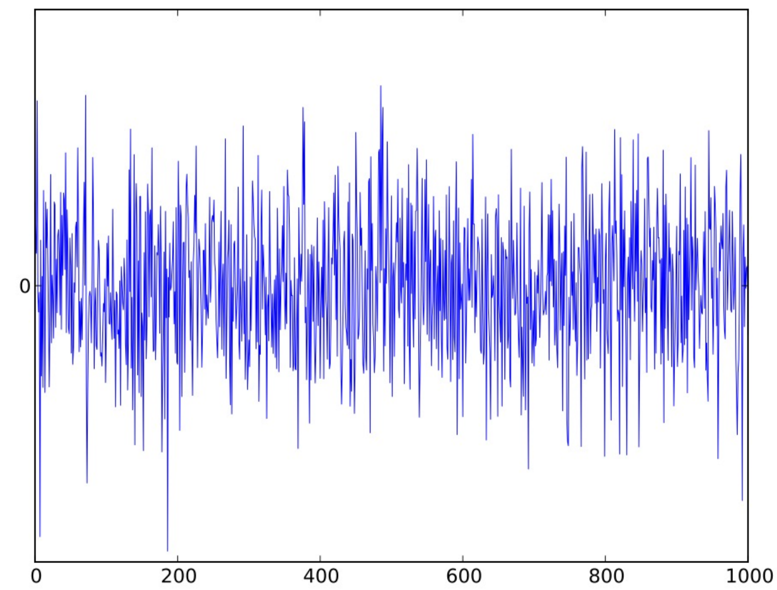
- Sound produced by warm air molecules without any external sound source = Gaussian noise
- Electrical signals: same.



Attribution Morn, https://commons.wikimedia.org/wiki/File:White_noise.svg

White Noise

- White Noise = noise in which each sample of the signal, x_n, is i.i.d.
- Why "white"? Because the Fourier transform, $X(\omega)$, is a zero-mean random variable whose variance is independent of frequency ("white")
- ***Gaussian White Noise***: x[n] are i.i.d. and Gaussian



Attribution Morn, https://commons.wikimedia.org/wiki/File:White_noise.svg

Contents

- Gaussian pdf; Central limit theorem, Brownian motion
- White Noise
- Vector of i.i.d. Gaussians
- Vector of Gaussians that are independent but not identical
- Facts about linear algebra
- Vector of Gaussians that are neither independent nor identical

Vector of Independent Gaussian Variables

Suppose we have a frame containing D samples from a Gaussian white noise process, $x_1, \ldots, x_D$. Let's stack them up to make a vector:

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_D \end{bmatrix}$$

This whole frame is random. In fact, we could say that $\vec{x}$ is a sample value for a Gaussian random vector called $\vec{X}$, whose elements are $X_1, \ldots, X_D$:

$$\vec{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_D \end{bmatrix}$$

Vector of Independent Gaussian Variables

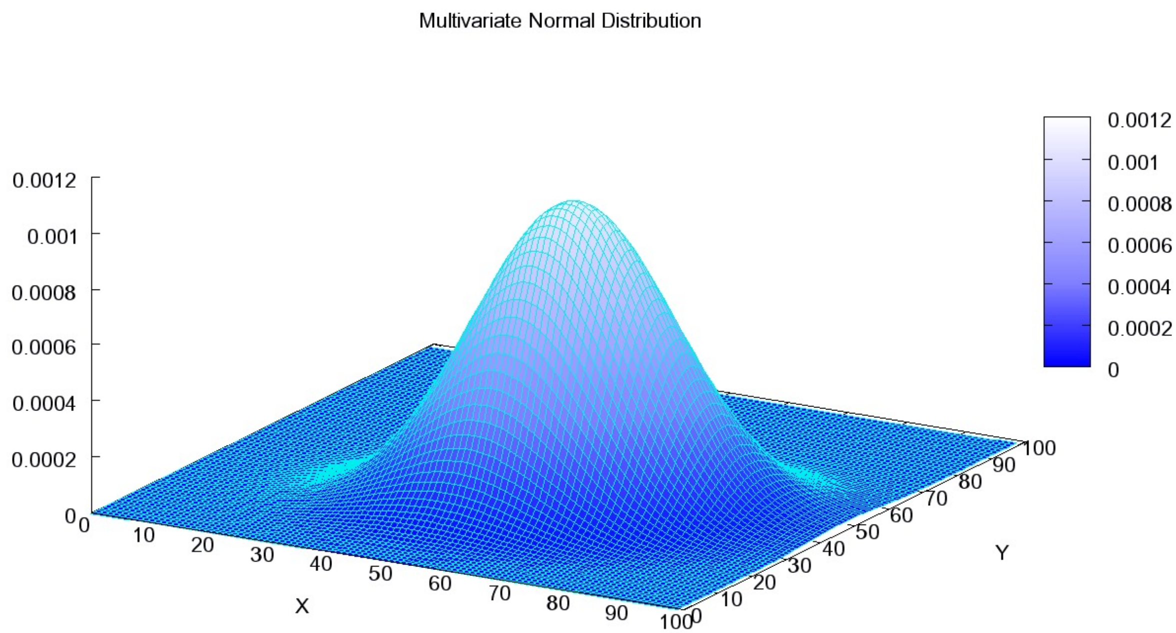
Suppose that the N samples are i.i.d., each one has the same mean, μ, and the same variance, σ^2. Then the pdf of this random vector is

$$p_{\vec{X}|Y}(\vec{x}|y) = \mathcal{N}(\vec{x}; \vec{\mu}, \sigma^2 I) = \prod_{n=1}^{D} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_n - \mu}{\sigma}\right)^2}$$

The class label, y, determines the mean and/or the variance of the Gaussian. For example, suppose that the label, y, is for a scene classifier. Traffic noise ($y =$"outside") has much higher energy (much higher σ^2) than the background noise in an office building ($y =$"inside"). So we assume that μ and σ^2 depend on y.

Vector of Independent Gaussian Variables

For example, here's an example from Wikipedia with mean of 50 and standard deviation of about 12.



Attribution: Piotrg, https://commons.wikimedia.org/wiki/File:Multivariate_Gaussian.png

Contents

- Gaussian pdf; Central limit theorem, Brownian motion
- White Noise
- Vector of i.i.d. Gaussians
- Vector of Gaussians that are independent but not identical
- Facts about linear algebra
- Vector of Gaussians that are neither independent nor identical

Independent Gaussians that aren't identically distributed

Suppose that the N samples are independent Gaussians that aren't identically distributed, i.e., X_d has mean μ_d and variance ${\sigma_d}^2$. The pdf of X_d is

$$p_{X_d|Y}(x_d|y) = \mathcal{N}(x_d; \mu_d, {\sigma_d}^2) = \frac{1}{\sqrt{2\pi{\sigma_d}^2}} e^{-\frac{1}{2}\left(\frac{x_d-\mu_d}{\sigma_d}\right)^2}$$

The pdf of this random vector is

$$p_{\vec{X}|Y}(\vec{x}|y) = \mathcal{N}(\vec{x}; \vec{\mu}, \Sigma) = \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi{\sigma_d}^2}} e^{-\frac{1}{2}\left(\frac{x_d-\mu_d}{\sigma_d}\right)^2}$$

Independent Gaussians that aren't identically distributed

Another useful form is:

$$\prod_{d=1}^{D} \frac{1}{\sqrt{2\pi{\sigma_d}^2}} e^{-\frac{1}{2}\left(\frac{x_d-\mu_d}{\sigma_d}\right)^2} = \frac{1}{(2\pi)^{D/2}\prod_{d=1}^{D}\sigma_d} e^{-\frac{1}{2}\sum_{d=1}^{D}\left(\frac{x_d-\mu_d}{\sigma_d}\right)^2}$$

Example

Suppose that $\mu_1 = 1, \mu_2 = -1, {\sigma_1}^2 = 1, {\sigma_2}^2 = 4.$ Then

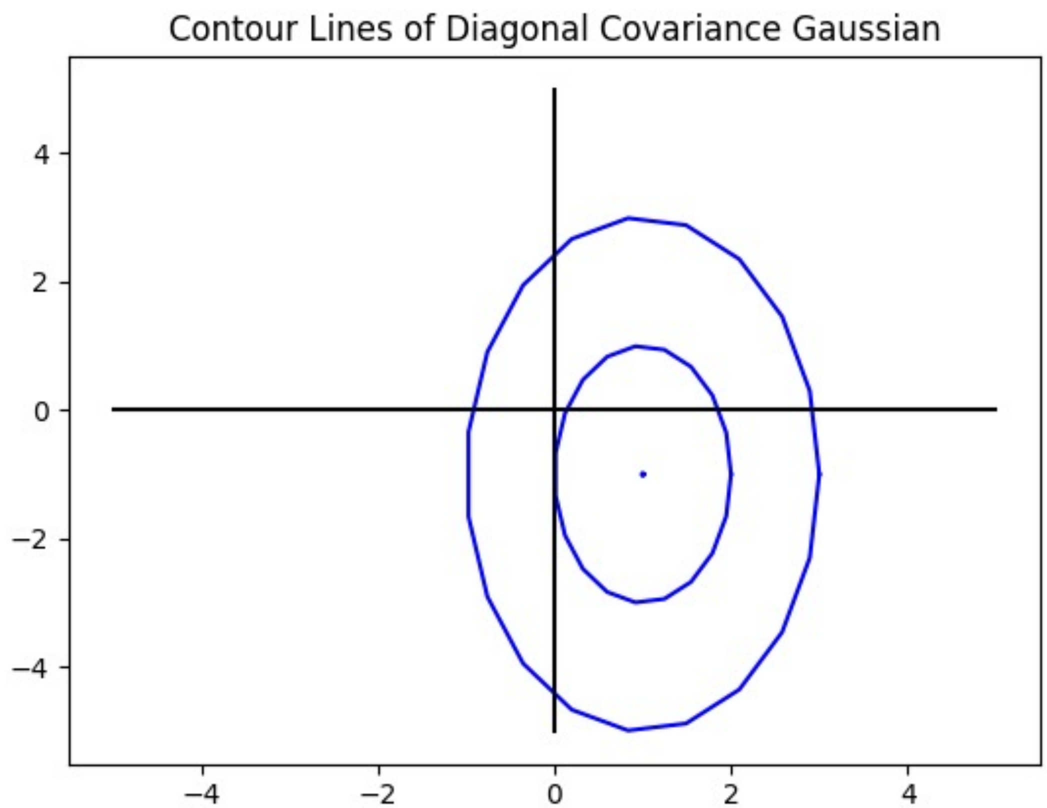
$$f_{\vec{X}}(\vec{x}) = \prod_{d=1}^{2} \frac{1}{\sqrt{2\pi{\sigma_d}^2}} e^{-\frac{1}{2}\left(\frac{x_d-\mu_d}{\sigma_d}\right)^2} = \frac{1}{4\pi} e^{-\frac{1}{2}\left(\left(\frac{x_1-1}{1}\right)^2+\left(\frac{x_2+1}{2}\right)^2\right)}$$

The pdf has its maximum value, $f_{\vec{X}}(\vec{x}) = \frac{1}{4\pi}$, at $\vec{x} = \vec{\mu} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

It drops to $\frac{1}{4\pi\sqrt{e}}$ at $\vec{x} = \begin{bmatrix} \mu_1 \pm \sigma_1 \\ \mu_2 \end{bmatrix}$ and at $\vec{x} = \begin{bmatrix} \mu_1 \\ \mu_2 \pm \sigma_2 \end{bmatrix}$.

It drops to $\frac{1}{4\pi e^2}$ at $\vec{x} = \begin{bmatrix} \mu_1 \pm 2\sigma_1 \\ \mu_2 \end{bmatrix}$ and at $\vec{x} = \begin{bmatrix} \mu_1 \\ \mu_2 \pm 2\sigma_2 \end{bmatrix}$.

Example



Contour Lines of Diagonal Covariance Gaussian
4
2
0
−2
−4
−4
−2
0
2
4

Contents

- Gaussian pdf; Central limit theorem, Brownian motion
- White Noise
- Vector of i.i.d. Gaussians
- Vector of Gaussians that are independent but not identical
- Facts about linear algebra
- Vector of Gaussians that are neither independent nor identical

Facts about linear algebra #1: determinant of a diagonal matrix

Suppose that Σ is a diagonal matrix, with variances on the diagonal:

$$\Sigma = \begin{bmatrix} {\sigma_1}^2 & 0 & 0 \\ 0 & {\sigma_2}^2 & \dots \\ 0 & \dots & {\sigma_D}^2 \end{bmatrix}$$

Then the determinant is

$$|\Sigma| = \prod_{d=1}^{D} {\sigma_d}^2$$

So we can write the Gaussian pdf as

$$\frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}\sum_{d=1}^{D}\left(\frac{x_d-\mu_d}{\sigma_d}\right)^2} = \frac{1}{|2\pi\Sigma|^{1/2}} e^{-\frac{1}{2}\sum_{d=1}^{D}\left(\frac{x_d-\mu_d}{\sigma_d}\right)^2}$$

Facts about linear algebra #2: inner product

Suppose that

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_D \end{bmatrix} \text{ and } \vec{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_D \end{bmatrix}$$

Then

$$(\vec{x} - \vec{\mu})^T (\vec{x} - \vec{\mu}) = (x_1 - \mu_1)^2 + \cdots + (x_D - \mu_D)^2$$

Facts about linear algebra #3: inverse of a diagonal matrix

Suppose that Σ is a diagonal matrix, with variances on the diagonal:

$$\Sigma = \begin{bmatrix} {\sigma_1}^2 & 0 & 0 \\ 0 & {\sigma_2}^2 & \ldots \\ 0 & \ldots & {\sigma_D}^2 \end{bmatrix}$$

Then its inverse, Σ^{-1}, is

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{{\sigma_1}^2} & 0 & 0 \\ 0 & \frac{1}{{\sigma_2}^2} & \ldots \\ 0 & \ldots & \frac{1}{{\sigma_D}^2} \end{bmatrix}$$

Facts about linear algebra #4: squared Mahalanobis distance with a diagonal covariance matrix

Suppose that all of the things on the previous slides are true.

Then the squared Mahalanobis distance is

$$d_{\Sigma}^{2}(\vec{x},\vec{\mu}) = (\vec{x}-\vec{\mu})^{T}\Sigma^{-1}(\vec{x}-\vec{\mu}) =$$

$$[x_1-\mu_1,\ldots,x_D-\mu_D]\begin{bmatrix} \frac{1}{{\sigma_1}^2} & 0 & 0 \\ 0 & \frac{1}{{\sigma_2}^2} & \ldots \\ 0 & \ldots & \frac{1}{{\sigma_D}^2} \end{bmatrix}\begin{bmatrix} x_1-\mu_1 \\ \vdots \\ x_D-\mu_D \end{bmatrix}$$

$$= \frac{(x_1-\mu_1)^2}{{\sigma_1}^2} + \cdots + \frac{(x_D-\mu_D)^2}{{\sigma_D}^2}$$

Mahalanobis form of the multivariate Gaussian, independent dimensions

So we can write the multivariate Gaussian as

$$p_{\vec{X}|Y}(\vec{x}|y) = \mathcal{N}(\vec{x}; \vec{\mu}, \Sigma) = \frac{1}{|2\pi\Sigma|^{1/2}}\, e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})}$$

$$p_{\vec{X}|Y}(\vec{x}|y) = \mathcal{N}(\vec{x}; \vec{\mu}, \Sigma) = \frac{1}{|2\pi\Sigma|^{1/2}}\, e^{-\frac{1}{2}d_\Sigma^2(\vec{x}-\vec{\mu})}$$

Facts about linear algebra #5: ellipses

The formula

$$1 = (\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu})$$

… or equivalently

$$1 = \frac{(x_1 - \mu_1)^2}{{\sigma_1}^2} + \cdots + \frac{(x_D - \mu_D)^2}{{\sigma_D}^2}$$

… is the formula for an ellipsoid (an ellipse in two dimensions; a football shaped object in three dimensions; etc.). The ellipse is centered at the point $\vec{\mu}$, and it has a volume proportional to $|\Sigma|$. (In 2D the area of an ellipse is $\pi|\Sigma|^{1/2}$, in 3D it's $\frac{4}{3}\pi|\Sigma|^{1/2}$, etc.)

Gaussian contour plots = ellipses

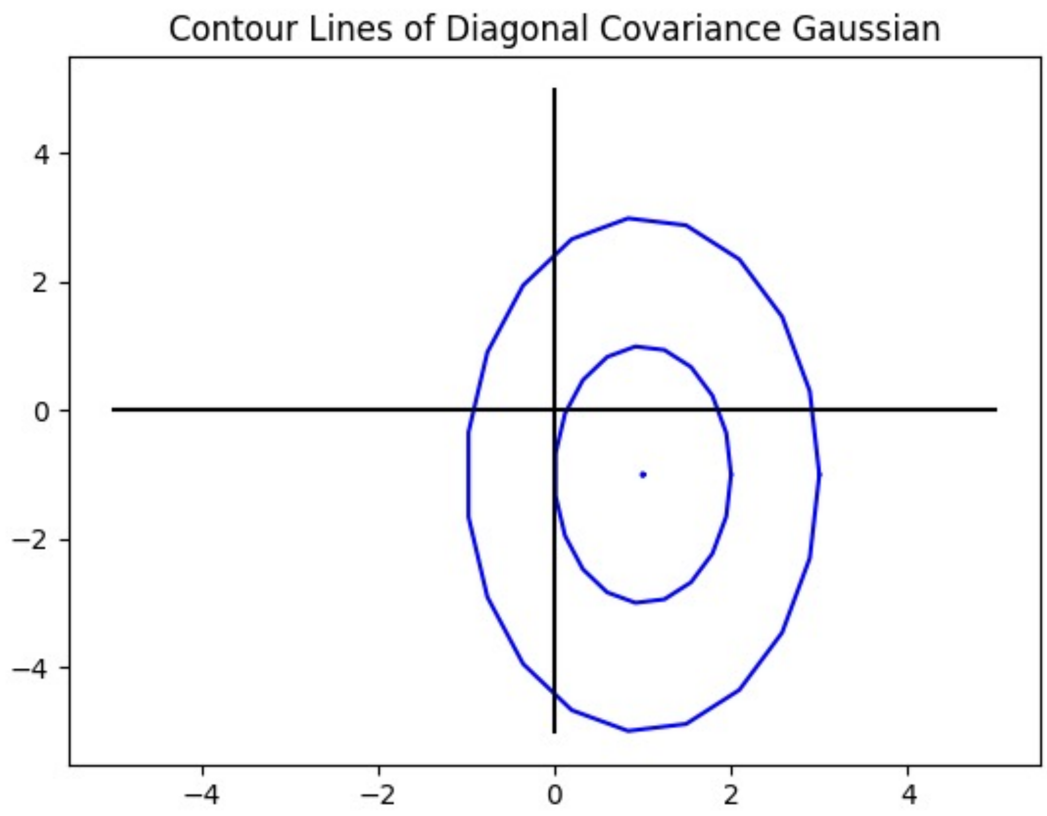
$$c = (\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu})$$

… is equivalent to

$$p_{\vec{X}|Y}(\vec{x}|y) = \frac{1}{|2\pi\Sigma|^{1/2}} e^{-\frac{1}{2}c}$$

Therefore the contour plot of a Gaussian pdf --- the curves of constant $f_{\vec{X}}(\vec{x})$ --- are ellipses. If Σ is diagonal, the main axes of the ellipse are parallel to the x_1, x_2, etc. axes. If Σ is NOT diagonal, the main axes of the ellipse are tilted.

Example


Contour Lines of Diagonal Covariance Gaussian
4
2
0
−2
−4
−4
−2
0
2
4

Contents

- Gaussian pdf; Central limit theorem, Brownian motion
- White Noise
- Vector of i.i.d. Gaussians
- Vector of Gaussians that are independent but not identical
- Facts about linear algebra
- Vector of Gaussians that are neither independent nor identical

Mahalanobis form of the multivariate Gaussian, dependent dimensions

If the dimensions are dependent, and jointly Gaussian, then we can still write the multivariate Gaussian as

$$p_{\vec{X}|Y}(\vec{x}|y) = \mathcal{N}(\vec{x}; \vec{\mu}, \Sigma) = \frac{1}{|2\pi\Sigma|^{1/2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})}$$

Example

Suppose that x_1 and x_2 are linearly correlated Gaussians with means 1 and -1, respectively, and with variances 1 and 4, and covariance 1.

$$\vec{\mu} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Remember the definitions of variance and covariance:

$$\sigma_1{}^2 = E[(x_1 - \mu_1)^2] = 1$$
$$\sigma_2{}^2 = E[(x_2 - \mu_2)^2] = 4$$
$$\sigma_{12} = \sigma_{21} = E[(x_1 - \mu_1)(x_2 - \mu_2)] = 1$$

$$\Sigma = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}$$

Determinant and inverse of a 2x2 matrix

You should know the determinant and inverse of a 2x2 matrix. If

$$\Sigma = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Then $|\Sigma| = ad - bc$ and

$$\Sigma^{-1} = \frac{1}{|\Sigma|}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

You should be able to verify the inverse, for yourself, by multiplying $\Sigma\Sigma^{-1}$ and discovering that the result is the identity matrix.

Example

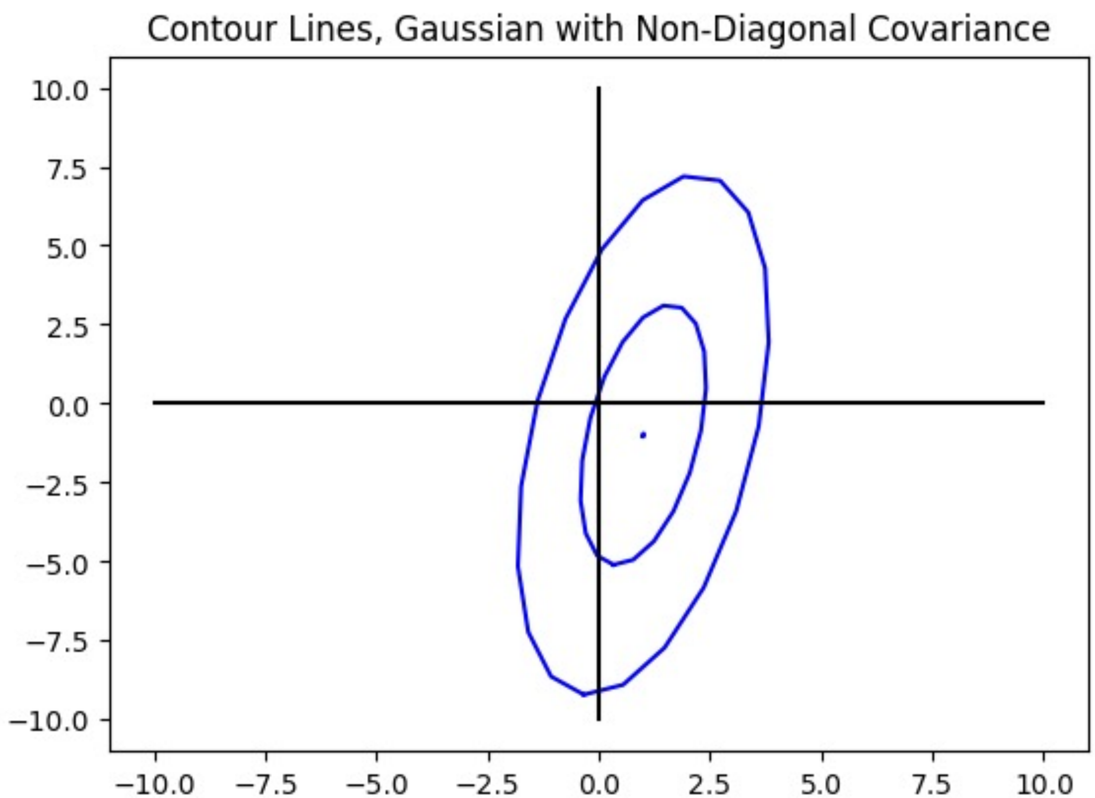
Therefore the contour lines of this Gaussian are ellipses centered at

$$\vec{\mu} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

The contour lines are ellipses that satisfy this equation. Each different value of c gives a different ellipse:

$$c = \frac{4}{3}(x_1 - 1)^2 + \frac{1}{3}(x_2 + 1)^2 - \frac{1}{3}(x_1 - 1)(x_2 + 1)$$

Example



Contour Lines, Gaussian with Non-Diagonal Covariance
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0

Conclusion: Summary of Today's Lecture

$$p_{\vec{X}|Y}(\vec{x}|y) = \mathcal{N}(\vec{x}; \vec{\mu}, \Sigma) = \frac{1}{|2\pi\Sigma|^{1/2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})}$$