SPR Day 6

Multivariate Gaussian Distribution

Reading: Bishop PRML Sec 2.3

intro

Sec 2.3.1 Conditionals are Gaussian

Sec 2.3.2 Marginals are Gaussian

Sec 2.3.3 Bayes for Gaussians

Sec 2.3.4 ML estimators

Sec 2.3.5 skim for intuition

Sec 2.3.6 skim for intuition from Fig 2.12

Outline:

1. Multivariate Gaussian PDF

2. Properties of covariance matrices

3. Marginal of a G is <u>G</u>aussian

4. Conditional of a G is <u>G</u>aussian

5. Linear Gaussian <u>joint</u> is <u>G</u>aussian

Joint Distribution of Two Independent Gaussians

Consider:
$$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$$
$$X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$$

X_1 is <u>indep.</u> of X_2

this implies

$\mathrm{Cov}[X_1, X_2] = 0$

Careful! This is a one-way implication.
$\mathrm{Cov}[A,B]=0$ does not mean
A is indep. of B.

Can write <u>joint</u> distribution's PDF as:

$$p(x_1, x_2) = \mathrm{NormPDF}(x_1 \mid \mu_1, \sigma_1^2)\ \mathrm{NormPDF}(x_2 \mid \mu_2, \sigma_2^2)$$

$$= c(\mu_1, \sigma_1^2)\, c(\mu_2, \sigma_2^2)\ e^{-\frac{1}{2}\frac{1}{\sigma_1^2}(x_1-\mu_1)^2}\ e^{-\frac{1}{2}\frac{1}{\sigma_2^2}(x_2-\mu_2)^2}$$

$$= \text{const} \cdot e^{-\frac{1}{2}(x-\mu)^T \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix} (x-\mu)}$$

where $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$

$$= \underbrace{\text{const}}_{\frac{1}{(2\pi)^{1/2}} \frac{1}{(2\pi)^{1/2}} \frac{1}{\sigma_1} \frac{1}{\sigma_2}} \cdot \exp\left\{ -\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right\}$$

$$= \frac{1}{(2\pi)^{D/2}} \frac{1}{|\det \Sigma|^{1/2}}$$

since $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix}$
$= ad - bc$
$= \sigma_1^2 \sigma_2^2$

Multivariate Gaussian PDF

Let X be a D-dim. random variable

$$X = [X_1, X_2, \cdots X_D]^T$$ (column vector with D entries)

Sample space is $\mathbb{R}^D$

Assume X has a multivariate Gaussian distribution

$$X \sim MVN(\mu, \Sigma)$$

Parameters:

"mean vector" $\mu \in \mathbb{R}^D$

"covariance" Σ is $D \times D$ — symmetric and positive definite

PDF function:

determinant of matrix Σ (scalar)

inverse of matrix Σ ($D \times D$)

$$\mathrm{MVNormPDF}(x \mid \mu, \Sigma) = \underbrace{\frac{1}{(2\pi)^{D/2}} \frac{1}{(\det \Sigma)^{1/2}}}_{c(\mu, \Sigma)} \underbrace{e^{-\frac{1}{2}\underbrace{(x-\mu)^T \Sigma^{-1} (x-\mu)}_{\text{Mahalanobis distance}}}}_{f(x, \mu, \Sigma)}$$

dimension check

$c(\mu, \Sigma) = \text{scalar} \cdot \text{scalar}$

$$f(x, \mu, \Sigma) = e^{-\frac{1}{2} \overbrace{\underbrace{(x-\mu)^T}_{1 \times D} \underbrace{\Sigma^{-1}}_{D \times D} \underbrace{(x-\mu)}_{D \times 1}}^{1 \times 1 \text{ scalar!}}}$$

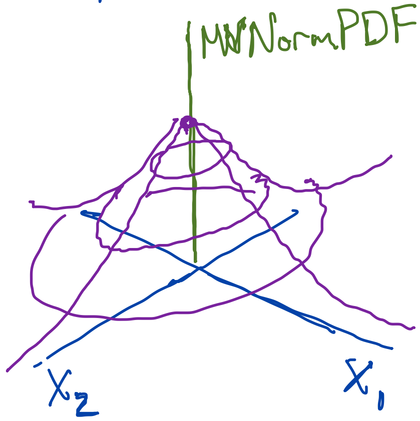
is it a valid PDF?

need: $\det \Sigma > 0$. Always true if Σ is positive def.

need: $\int e^{-\frac{1}{2} d(x)} dx$ to be finite. True if Σ is positive def. so this is a unimodal function.

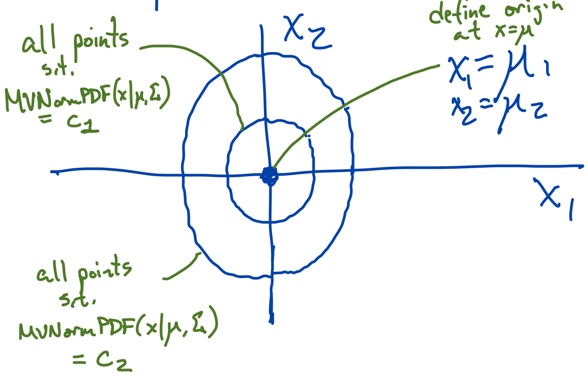
Basic Facts and Visuals

3D plot



unimodal "bump"
peak at $x = \mu$
smoothly decaying away as dist from μ increases

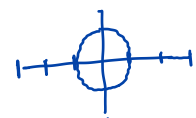
2D plot contours



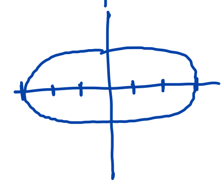
Pick several "levels" with fixed PDF value c_1, c_2

Turns out, all level sets look like ellipses in 2D, with center $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$

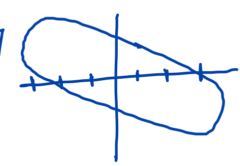
$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ perfect circle



$\Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$ stretch x_1



$\Sigma = \begin{bmatrix} 3 & -0.5 \\ -0.5 & 1 \end{bmatrix}$ stretch x_1 and rotate



Given: $x \sim \mathcal{N}(\mu, \Sigma)$:

Mode (peak) is at $x = \mu$

mean of X is at $\mathbb{E}[x] = \mu$

covariance of X is $\text{Cov}[x] = \Sigma$

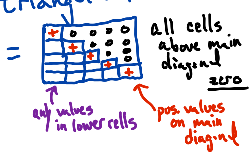
expected value of xx^T is $\mathbb{E}[xx^T] = \Sigma + \mu\mu^T$

Covariance Matrix Properties

Assume parameter Σ is <u>symmetric</u> and <u>positive definite</u>.

Then we know:

- Σ is invertible: There exists a $D \times D$ matrix Σ^{-1} s.t. $\Sigma \Sigma^{-1} = I$
  and Σ^{-1} is <u>also</u> pos. def.
- Σ has all positive diagonal entries
- Σ has positive determinant $\det \Sigma > 0$
- Σ has a unique Cholesky factorization. $\Sigma = L L^T$ for some $D \times D$ lower triangular matrix

$L =$ 

- Σ has all positive eigenvalues.

Let Σ be decomposed into $\lambda_d \geq 0 \ \forall d$

D eigenvalues $\lambda_1, \lambda_2, \cdots \lambda_D$

D eigenvectors $u_1, u_2, \cdots u_D$ each $u_d \in \mathbb{R}^D$

These form an orthonormal <u>basis</u>

$u_d^T u_d = 1 \ \forall d$

$u_d^T u_e = 0 \ \forall d \neq e$

Each u_d is orthogonal to all other eigenvectors and has magnitude 1 (its a unit vector)

We can write:

$$\Sigma = \sum_{d=1}^{D} \lambda_d \underbrace{u_d u_d^T}_{D \times D \text{ outer product}}$$

and it turns out the <u>inverse</u> is:

$$\Sigma^{-1} = \sum_{d=1}^{D} \frac{1}{\lambda_d} u_d u_d^T$$

How does this eigen decomposition help?

Define $U = \begin{bmatrix} — u_1 — \\ \vdots \\ — u_D — \end{bmatrix}$ as $D \times D$ matrix of stacked eigenvectors.

Mahalanobis distance becomes

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

$$= (x-\mu)^T U^T \Lambda U (x-\mu)$$

$\Lambda = \begin{bmatrix} \frac{1}{\lambda_1} & & \\ & \ddots & \\ & & \frac{1}{\lambda_D} \end{bmatrix}$ diagonal matrix of inverted eigenvalues

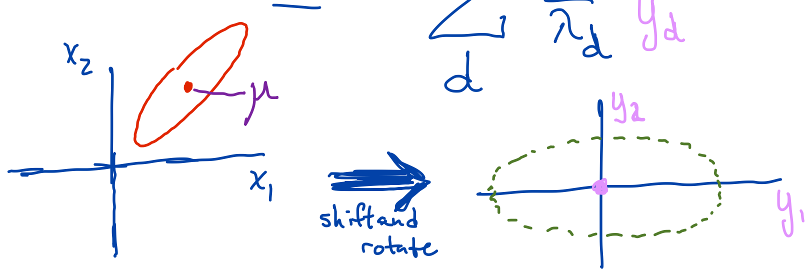
Because U is square with determinant 1 and $U^T = U^{-1}$, it is a rotation matrix.

Change variables: $y(x) = U(x-\mu)$ shift by μ, rotate by U

$$\mathrm{dist} = y(x)^T \Lambda \, y(x)$$

$$= \sum_d \frac{1}{\lambda_d} y_d^2$$

equation for an axis-aligned ellipse



Thus, all x values with same distance (same PDF density value) live on an elliptical contour.

Joint Gaussian Distribution and possible partitions

Suppose we have D-dimensional r.v. X s.t. $X \sim MVN(\mu, \Sigma)$. Assume $\mu \in \mathbb{R}^D$, $\Sigma \in$ sym. pos. def. $D \times D$. Both known.

Given any reordering of dimension indices

π is permutation of $\{1, 2, \cdots D\}$

then,
$$X_\pi = \begin{bmatrix} x_{\pi(1)} \\ \vdots \\ x_{\pi(D)} \end{bmatrix} \sim MVN\left(\begin{bmatrix} \mu_{\pi(1)} \\ \vdots \\ \mu_{\pi(D)} \end{bmatrix}, \begin{bmatrix} \Sigma_{\pi(1)\pi(1)} & \cdots & \Sigma_{\pi(1)\pi(D)} \\ \vdots & \ddots & \vdots \\ \Sigma_{\pi(D)\pi(1)} & & \Sigma_{\pi(D)\pi(D)} \end{bmatrix} \right)$$

Consider any partition of indices

$$\{\underbrace{1, 2, 3, \cdots}_{A} \underbrace{D-2, D-1, D}_{B}\}$$

first D^A dims belong to A — remaining dims belong to B

then
$$\begin{bmatrix} X_A \\ X_B \end{bmatrix} \sim MVN\left(\begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix}, \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{AB}^T & \Sigma_{BB} \end{bmatrix} \right)$$

μ is $D \times 1$ $\quad$ Σ is $D \times D$

Σ_{AA} is square matrix $D^A \times D^A$

Σ_{BB} is square matrix $D^B \times D^B$

Σ_{AB} is rectangular $D^A \times D^B$

Marginals of a Joint Gaussian are <u>Gaussian</u>

Given joint distribution over x_A and x_B:

$$\begin{bmatrix} x_A \\ x_B \end{bmatrix} \sim MVN\left(\begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix}, \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{AB}^T & \Sigma_{BB} \end{bmatrix} \right)$$

μ is $D \times 1$ Σ is $D \times D$

Σ_{AA} is square matrix $D^A \times D^A$

Σ_{BB} is square matrix $D^B \times D^B$

Σ_{AB} is <u>rectangular</u> $D^A \times D^B$

We have the marginal of x_A is:

$$p(x_A) = \int p(x_A, x_B)\, dx_B$$

$$= \text{MVNormPDF}(x_A \mid \mu_A, \Sigma_{AA})$$

By symmetry

$$p(x_B) = \text{MVNormPDF}(x_B \mid \mu_B, \Sigma_{BB})$$

Conditionals of a Joint Gaussian are Gaussian

Given joint distribution over x_A, x_B:

$$\begin{bmatrix} x_A \\ x_B \end{bmatrix} \sim MVN\left(\begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix}, \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{AB}^T & \Sigma_{BB} \end{bmatrix} \right)$$

μ is $D \times 1$ $\qquad$ Σ is $D \times D$

Σ_{AA} is square matrix $D^A \times D^A$

Σ_{BB} is square matrix $D^B \times D^B$

Σ_{AB} is rectangular $D^A \times D^B$

Recall that Σ^{-1} exists. Define $\Delta = \Sigma^{-1}$

with partitions $\Delta = \begin{bmatrix} \Delta_{AA} & \Delta_{AB} \\ \Delta_{AB}^T & \Delta_{BB} \end{bmatrix}$

$D \times D$ overall

Δ_{AA} is square $D^A \times D^A$

Δ_{BB} is square $D^B \times D^B$

Δ_{AB} is rectangle $D^A \times D^B$

Define the conditional density of x_A given $x_B = m_B$

$$p(x_A \mid x_B = m_B) = \frac{p(x_A, x_B = m_B)}{p(x_B = m_B)}$$

known value of x_B $\qquad$ mean of x_B

$$= MVNormPDF\left(x_A \mid \mu_A - \Delta_{AA}^{-1} \Delta_{AB} (m_B - \mu_B), \Delta_{AA}^{-1}\right)$$

dim. check $\quad A \times 1 \qquad (A \times A)\ (A \times B)(B \times 1) \qquad A \times A$

$\qquad\qquad A \times 1$

verified! mean is $A \times 1$

cov. is $A \times A$

Linear Gaussian Model

Consider a model with two random variables:

(1) $\underset{}{x} \sim MVN(\underset{S\times 1}{\mu}, \underset{S\times S}{\Lambda}^{-1})$ $\quad x \in \mathbb{R}^S$

(2) $y \mid x \sim MVN(\underset{T\times S}{A}\underset{S\times 1}{x} + \underset{T\times 1}{b}, \underset{T\times T}{L}^{-1})$ $\quad y \in \mathbb{R}^T$

What is the joint distribution?

Write log pdf out and simplify

$$\log p(x,y) = \log p(x) + \log p(y|x)$$

$$= \underset{\text{wrt } x,y}{\text{const}} - \frac{1}{2}(x-\mu)^T \Lambda (x-\mu) - \frac{1}{2}(y-Ax-b)^T L (y-Ax-b)$$

$$= \underset{\text{wrt } x,y}{\text{const}} - \frac{1}{2} x^T \Lambda x - \frac{1}{2} x^T A^T L A x - \frac{1}{2}(-2) y^T L A x - \frac{1}{2} y^T L y$$

$$- \frac{1}{2}(-2) \underbrace{x^T \Lambda \mu}_{\text{order } 1} - \frac{1}{2}(+2) \underbrace{x^T A^T L b}_{\text{order } 1} - \frac{1}{2}(-2) \underbrace{y^T L b}_{\text{order } 1}$$

$$= \underset{\text{wrt } x,y}{\text{const}} \underbrace{- \frac{1}{2} \begin{pmatrix} x \\ y \end{pmatrix}^T \begin{bmatrix} \Lambda + A^T L A & -A^T L \\ -LA & L \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix}}_{\text{order } 2} + \underbrace{\begin{pmatrix} \Lambda\mu - A^T L b \\ L b \end{pmatrix}^T \begin{pmatrix} x \\ y \end{pmatrix}}_{\text{order } 1}$$

Continued....

$$= \underset{\text{wrt } x,y}{\text{const}} - \frac{1}{2}\begin{pmatrix}x\\y\end{pmatrix}^T P \begin{pmatrix}x\\y\end{pmatrix} + \begin{pmatrix}x\\y\end{pmatrix}^T \left(P m\right)$$

where

$$P = \begin{bmatrix} \Delta + A^T L A & -A^T L \\ -LA & L \end{bmatrix}$$

shape of P
$(S+T) \times (S+T)$

$$m = P^{-1}\begin{bmatrix} \Delta\mu - A^T L b \\ L b \end{bmatrix}$$

shape of m
$S+T \times 1$

thus:

$$\log p(x,y) = \text{const} - \frac{1}{2}\left(\begin{bmatrix}x\\y\end{bmatrix} - m\right)^T P \left(\begin{bmatrix}x\\y\end{bmatrix} - m\right)$$

Recognise this is a Multivariate Gaussian over $\mathbb{R}^{S+T}$
with mean m
covariance P^{-1}