Day 6
Multivariate Gaussians

Mean and Covariance
PDF function
Understanding covariance matrices
- symmetric
- positive definite
- determinants

Graphing in 2D
- contour plots
- see notebook online!

Multivariate Gaussian

Let X be a Multivariate Gaussian Rand. Var.
Sample space is: $X \in \mathbb{R}^D$ (D dimensions)
$X = [x_1, x_2, \cdots x_D]^T$ as a column vector

Parameters

$\mu \in \mathbb{R}^D$ is the D-dim mean vector

Σ is a $D \times D$ covariance matrix

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1D} \\ \Sigma_{21} & \Sigma_{22} & & \\ \vdots & & \ddots & \\ \Sigma_{D1} & \cdots & & \Sigma_{DD} \end{bmatrix}$$

must be symmetric and positive definite

we'll define this in a bit...

PDF function

$$\text{MultivarNormPDF}(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp\Big[-\tfrac{1}{2}\underbrace{(x-\mu)^T}_{1\times D}\underbrace{\Sigma^{-1}}_{D\times D}\underbrace{(x-\mu)}_{D\times 1}\Big]$$

dimension analysis: Scalar · scalar · exp[scalar]

is it a valid pdf?

$\text{pdf}(x) > 0 \ \forall x$ $\quad \frac{1}{(2\pi)^{\frac{D}{2}}} > 0 \quad \frac{1}{|\Sigma|} > 0$ when? $\quad e^{\text{scalar}} > 0$ when?

Special case: $D = 2$ and $\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$

What is pdf of this special case?

$$\text{MVNormPDF}\left(\begin{matrix} x_1 \\ x_2 \end{matrix} \middle| \begin{matrix} \mu_1 \\ \mu_2 \end{matrix}, \begin{matrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{matrix} \right)$$

$$\propto \text{const} \cdot e^{-\frac{1}{2}\left[\frac{1}{\sigma_1^2}(x_1-\mu_1)^2 + \frac{1}{\sigma_2^2}(x_2-\mu_2)^2\right]}$$

$$\propto \text{const} \cdot \text{NormPDF}(x_1 | \mu_1, \sigma_1^2) \cdot \text{NormPDF}(x_2 | \mu_2, \sigma_2^2)$$

product of in<u>dependent</u> marginals

Any time we have a covariance matrix where <u>only</u> non-zero entries are on diagonal, this means we are making independence assumption

x_1 indep of $x_2, \dots x_D$

x_2 indep of $x_3, \dots x_D$

etc.

Expectations: When $x \sim \text{MVNorm}(\mu, \Sigma)$

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

$$\mathbb{E}[x x^T] = \mu\mu^T + \Sigma$$

$$\text{Cov}[x_i, x_j] = \Sigma_{ij}$$

In general ~~for~~ any rand. var that is a vector:

$$\text{Cov}[\vec{x}] = \mathbb{E}\left[(x - \mathbb{E}[x])(x - \mathbb{E}[x])^T\right]$$

$D \times D$ matrix

where

$$\text{Cov}[x_i, x_j] = \mathbb{E}\left[(x_i - \mathbb{E}[x_i])(x_j - \mathbb{E}[x_j])\right]$$

Linear Algebra Review:
Symmetric and Positive Definite Matrix

Let matrix A be a $D \times D$ square matrix

We say A is symmetric iff $A = A^T$

example: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix}$ are symmetric

but $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ is NOT

Property: Symmetric matrix has all real eigenvalues

We say A is positive definite
if for all vectors $x \in \mathbb{R}^D$ not equal to all zero vector,

$$x^T A x > 0$$ strictly greater

weaker condition: positive semi-definite
allows equality

$$x^T A x \geq 0$$

examples:

$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & -1.1 \\ -1.4 & 4 \end{bmatrix}$ are positive definite

Remember: x^TAx is really:

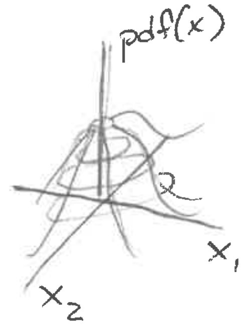
$$[x_1 \; x_2] \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + a_{22}x_2^2$$

Properties: If A is positive definite then:
- it is invertible
- all columns are linearly indep.
- all eigenvalues are positive

Graphical Understanding of 2-dimensional MV Gaussian

PDF: $\quad \text{const}(\mu, \Sigma) \cdot e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$

in 3D:



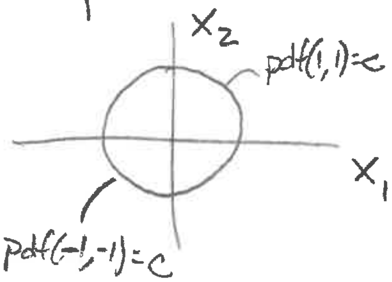
unimodal bump
smoothly
decaying away from $x=\mu$

μ determines location

Σ determines "width" and "correlation"

contour plot: contours look like ellipses

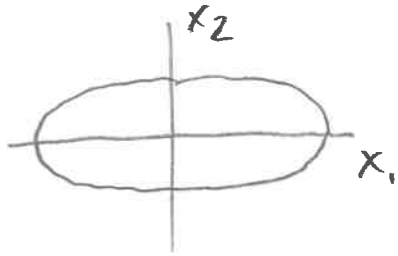
$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$



look at curves in $x_1 \times x_2$ plane
with constant pdf value

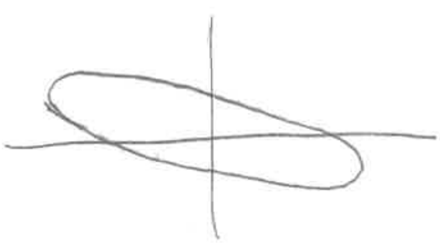
perfect circle w/ identity covariance

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



"stretch" along x_1 axis

$\Sigma = \begin{bmatrix} 3 & -0.5 \\ -0.5 & 1 \end{bmatrix}$



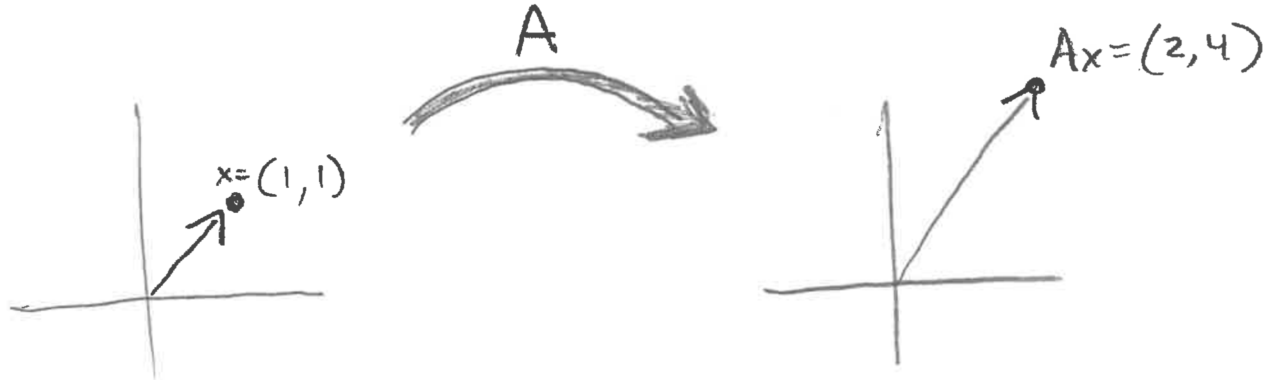
"rotate"
with slight
negative correlation

$\Sigma = \begin{bmatrix} 3 & +0.5 \\ +0.5 & 1 \end{bmatrix}$

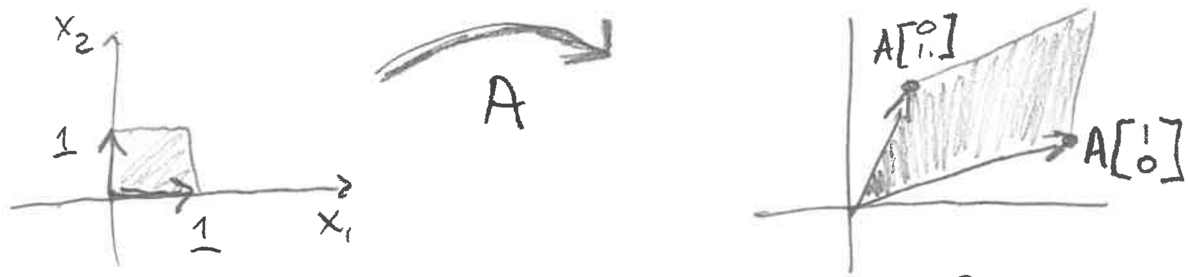
"rotate"
with slight
positive
correlation

Linear Algebra Review: Determinant

Think of square matrix A as mapping from $\mathbb{R}^D \to \mathbb{R}^D$



Question:
How much area/volume change after this transform?



area of this parallelogram is det(A)

In 2D: $\det(A) = a_{11}a_{22} - a_{12}a_{21}$

In General: $\det(A) = \prod_{d=1}^{D} \lambda_d$ product of eigenvalues

Key Idea: determinants capture signed volume transform

det can be $-3, -0.5, 0, 0.5, 3.8, \ldots$

negative means reverse orientation

uninvertible contraction

Eigenvector / Eigenvalue representation

We assume Σ <u>must</u> be positive definite and symmetric

All equivalent statements:

- Σ is positive definite
- Σ is invertible <u>and</u> has all positive diagonal entries ($\exists$ a $D \times D$ matrix Σ^{-1} s.t. $\Sigma \Sigma^{-1} = I$)
- Σ has positive determinant <u>and</u> all positive diagonal entries
- Σ has all positive eigenvalues $\begin{cases} \lambda_1 > 0 \\ \vdots \\ \lambda_D > 0 \end{cases}$
- Σ has a unique cholesky factorization $\Sigma = L L^T$ where L is lower triangular
- inverse is positive definite

We can write Σ in terms of D eigenvectors $u_1, \dots u_D$ and D eigenvalues $\lambda_1 \dots \lambda_D$

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

where eigenvectors form an <u>orthonormal</u> <u>basis</u> $u_d^T u_d = 1 \ \forall d$, $u_d^T u_f = 0 \ \forall d \neq f$

Turns out that we can write

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

Proof sketch (not required for HW or Quiz)

$\Sigma \Sigma^{-1}$ must equal I

$$\left(\sum_d \lambda_d u_d u_d^T\right)\left(\sum_f \frac{1}{\lambda_f} \mu_f \mu_f^T\right)$$

$d = f$ same terms

$d \neq f$ cross terms are orthogonal

$$\sum_{d=f} \frac{\lambda_d}{\lambda_d} u_d (u_d^T u_d) u_d^T + \sum_{d \neq f} \frac{\lambda_d}{\lambda_f} u_d u_d^T u_f u_f^T$$

$\checkmark$ $\emptyset$ ortho good.

$$\sum_d 1 \cdot u_d u_d^T = I$$

basis: think about Ux $\quad x = a_1 u_1 + \dots a_D u_D$

$$= \sum_d u_d (u_d^T x) = \sum_d u_d a_d = x$$

From Joint Gaussian Distribution to Marginal + Conditional

We have vector $x = [x_1, x_2, x_3, \ldots x_{D-1}, x_D]$
in D dimensions

Consider any exclusive partition of indices into A and B

e.g. if $D = 4$, then $A = \{1\}$, $B = \{2, 3, 4\}$
or
$A = \{2, 3\}$, $B = \{1, 4\}$

Define $x \sim \mathcal{N}(\mu, \Sigma)$ assuming we've reordered dims so we count first thru A then B

this means:

$$\begin{bmatrix} x_A \\ x_B \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix}, \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix} \right)$$

D x 1 vector

D x D cov matrix

Section 2.3.1 + 2.3.2
Bishop has lots of math deriving the following:

What is the conditional $p(x_A \mid x_B)$?

What is the marginal $p(x_A)$?

$p(x_B)$?

You should understand the math, but we'll skip math and show results

From Joint Gaussian to Marginal + Conditional

$$\begin{bmatrix} x_A \\ \hline x_B \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_A \\ \hline \mu_B \end{bmatrix}, \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix} \right)$$

whole covariance matrix

Σ is $D \times D$

has an inverse

$$\Lambda = \Sigma^{-1}$$

$$= \begin{bmatrix} \Lambda_{AA} & \Lambda_{AB} \\ \Lambda_{BA} & \Lambda_{BB} \end{bmatrix}$$

Marginal

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

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

Conditional

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

$$= N\left(x_A \mid \mu_A - \Lambda_{AA}^{-1}\Lambda_{AB}(x_B - \mu_B), \Lambda_{AA}^{-1}\right)$$

dim check: $A{\times}1 - (A{\times}A)(A{\times}B)(B{\times}1 - B{\times}1)$, $A{\times}A$ ✓

$= A \times 1$ ✓