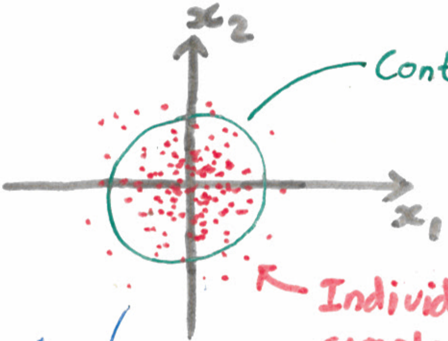
Multivariate Gaussians + Covariance

$$x_d \sim N(0,1) \Rightarrow p(\underline{x}) = \frac{1}{(2\pi)^{D/2}} e^{-\underline{x}^T\underline{x}/2}$$

Contours of $p(\underline{x})$ 'spherical', const. radius



Individual samples

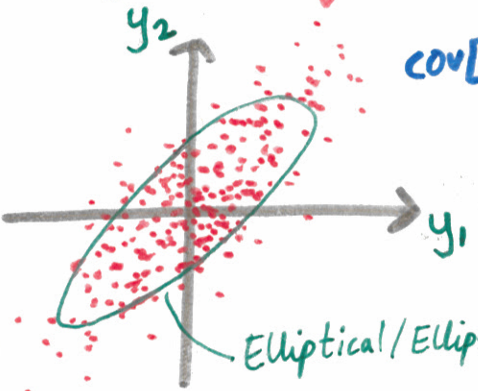
$$\text{cov}[\underline{x}] = \mathbb{I} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

DXD matrix

$y = A\underline{x}$

$$\text{cov}[\underline{x}] = \mathbb{E}[\underline{x}\,\underline{x}^T] - \mathbb{E}[\underline{x}]\,\mathbb{E}[\underline{x}]$$

$$\text{cov}[\underline{x}]_{ij} = \mathbb{E}[x_i x_j] - \mathbb{E}[x_i]\,\mathbb{E}[x_j]$$

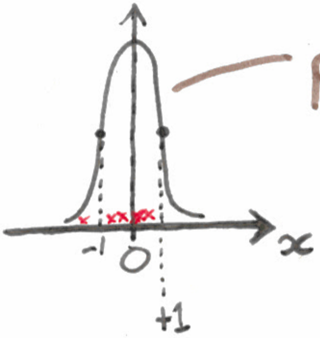


Elliptical / Ellipsoidal

$$\text{cov}[y] = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$$

$$= AA^T = \Sigma$$

Univariate Gaussian Reminder



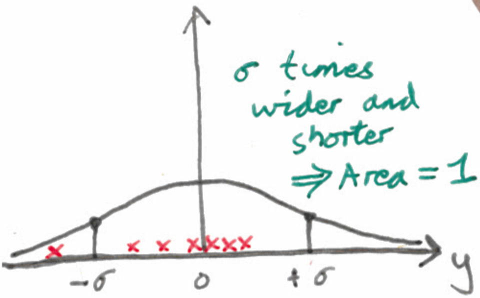
$\rightarrow$ $\approx 2/3$ samples $\leftarrow$

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2$$

Transform:

$$y = \sigma x, \quad x = \frac{y}{\sigma}$$



$\rightarrow$ $\approx 2/3$ samples $\leftarrow$

$$p(y) = \frac{1}{\underbrace{\sigma}_{\text{scaling}} \sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}}$$

Transforming the spherical dist

$\underline{y} = A\underline{x}$ stretch, rotation, reflect

$$\mathrm{cov}[\underline{y}] = \mathbb{E}\left[(A\underline{x})(A\underline{x})^T\right] - \cancel{A\underbrace{\mathbb{E}[\underline{x}]}_{0}\mathbb{E}[A\underline{x}]^T}$$

$$= \mathbb{E}\left[A\underline{x}\,\underline{x}^T A^T\right]$$

$$= A\underbrace{\mathbb{E}[\underline{x}\,\underline{x}^T]}_{I}A^T$$

$$\boxed{\mathrm{cov}[\underline{y}] = \Sigma = AA^T}$$

PDF of y, $N(y; 0, \Sigma)$

$$p(y) \propto e^{-(A^{-1}y)^T(A^{-1}y)/2} \qquad (\underline{x} = A^{-1}y)$$

$$\propto e^{-\frac{1}{2}y^T\underbrace{(A^{-T}A^{-1})}y}$$

If $\Sigma = AA^T$

$\Sigma^{-1} = A^{-T}A^{-1}$

$$p(y) = \frac{1}{|A|(2\pi)^{D/2}} e^{-\frac{1}{2}y^T\Sigma^{-1}y}$$

volume increases by $|A|$

$|A| = |\Sigma|^{1/2}$

$$|\Sigma| = |AA^T| = |A||A^T| = |A|^2$$

General Gaussian

$$\underline{z} = \underbrace{A\underline{x}}_{y} + \underline{\mu} \qquad y = \underline{z} - \underline{\mu}$$

$$\boxed{\begin{aligned} p(\underline{z}) &= \frac{1}{|\Sigma|^{1/2}(2\pi)^{D/2}} e^{-1/2(z-\mu)^T\Sigma^{-1}(z-\mu)} \\ &= N(\underline{z}; \underline{\mu}, \Sigma) \end{aligned}}$$

Iff $\Sigma = AA^T$, then it's positive semi-definite

P.S.D. $\underline{z}^T \Sigma \underline{z} \geqslant 0$ for all $\underline{z}$

Also $\underline{z}^T \Sigma^{-1} \underline{z} \geqslant 0$

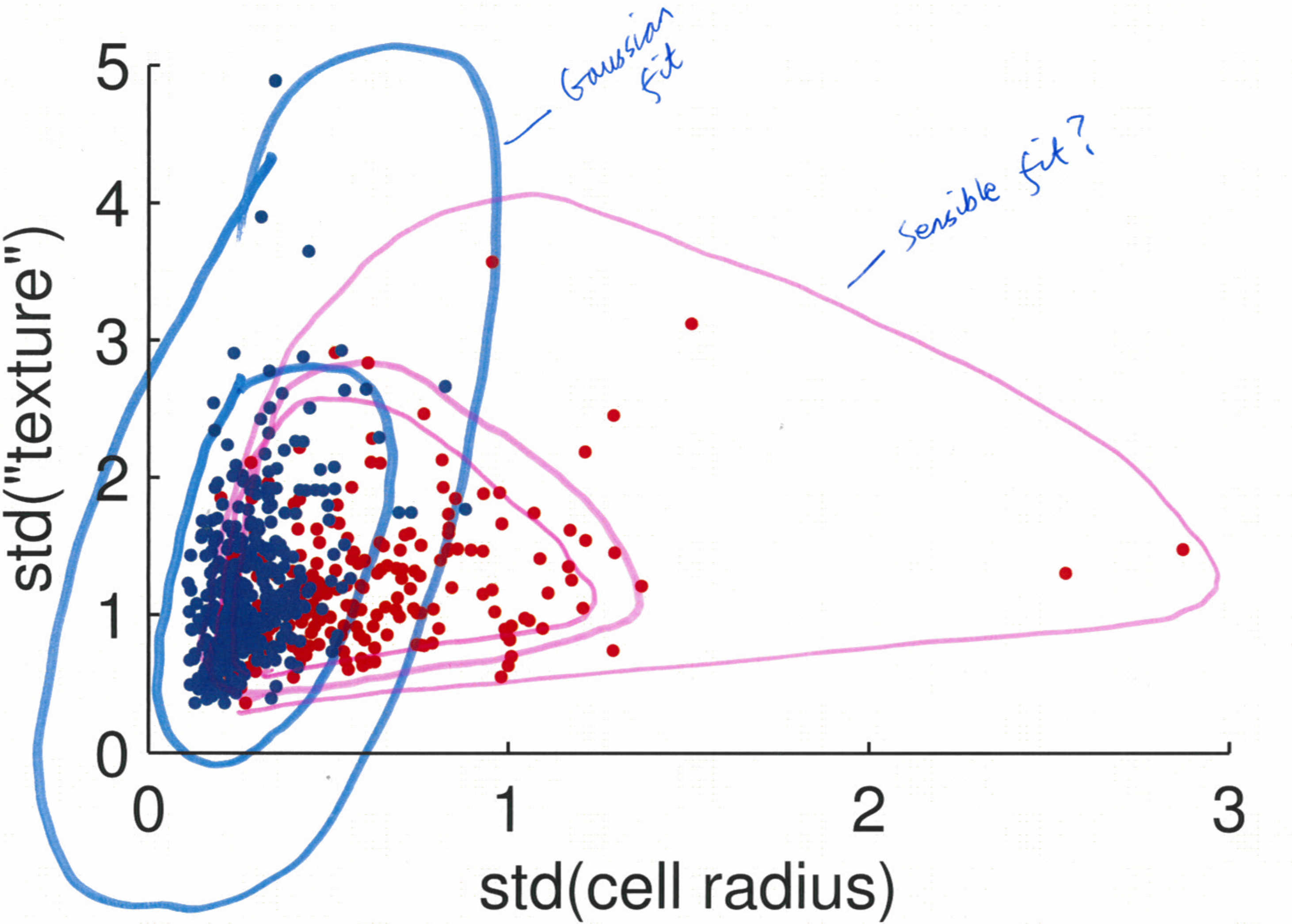
Example where semi definite

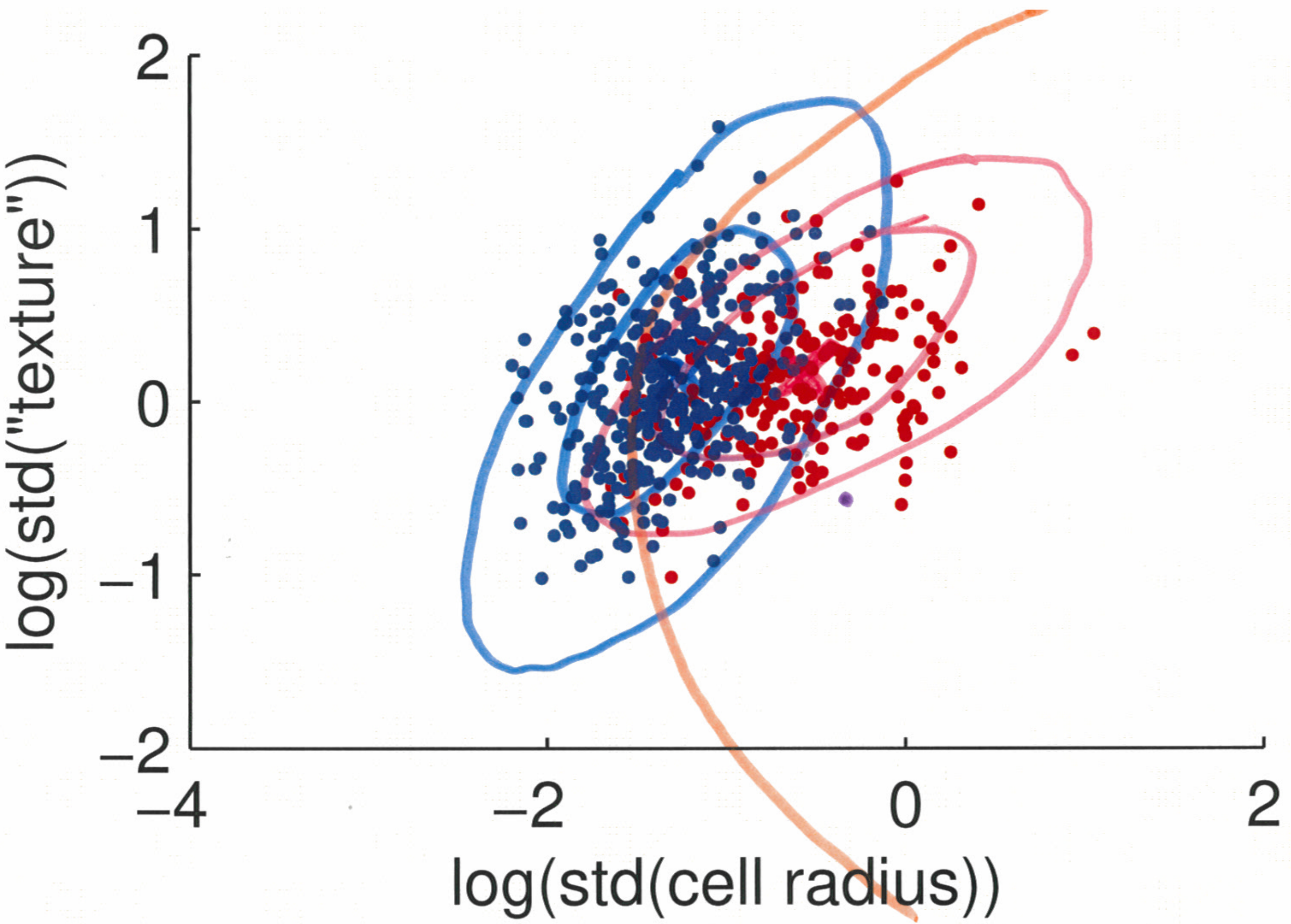
$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Positive Definite case:

$L = \text{chol}(\Sigma)$, $\Sigma = LL^T$

Other decompositions tutorial 2


std("texture")
0
1
2
3
4
5
std(cell radius)
0
1
2
3
Gaussian fit
Sensible fit?



log(std("texture"))
2
1
0
−1
−2
−4
−2
0
2
log(std(cell radius))

Fit Bayes Classifier

Model specifies

Dist over labels $p(y)$

Dist features for each class

$p(\underline{x} \mid y=k)$

$k = 1 \ldots K$

↑

classes

Could fit

$$p(y=1) = \pi_1 \approx \frac{\#\text{ in class } 1}{N}$$

Apply Bayes' Rule

$$p(y=k \mid \underline{x}) \propto p(\underline{x} \mid y=k)\, p(y=k)$$

$$s_k = N(\underline{x};\, \mu^{(k)}, \Sigma^{(k)})\, \pi_k$$

$$p(y=k \mid \underline{x}) = \frac{s_k}{\sum_{k'} s_{k'}}$$