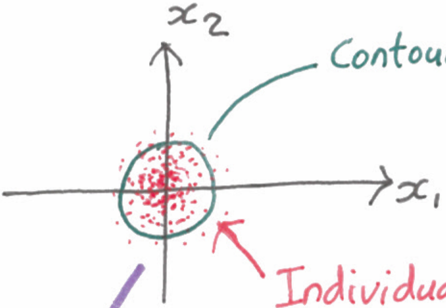
Multivariate Gaussians

$$x_d \sim \mathcal{N}(0,1) \Rightarrow p(\underline{x}) = \frac{1}{(2\pi)^{D/2}}\, e^{-\underbrace{\underline{x}^\top \underline{x}}_{\text{square radius}}/2}$$

Contours of $p(\underline{x})$

circular/spherical



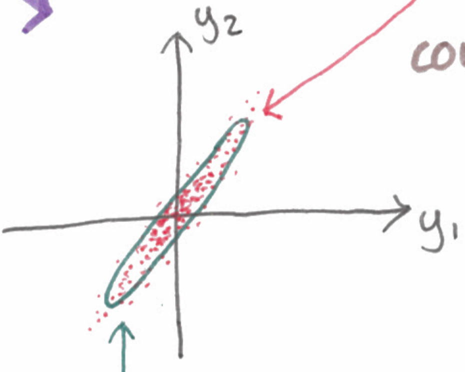
Individual samples $\underline{x}^{(n)}$ or $\underline{y}^{(n)}$

$$\underline{y}^{(n)} = A\,\underline{x}^{(n)}$$

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

$$\mathrm{cov}[\underline{y}] = \mathbb{E}[\underline{y}\underline{y}^\top] - \mathbb{E}[\underline{y}]\,\mathbb{E}[\underline{y}]^\top$$

$$= AA^\top = \underbrace{\Sigma}_{\text{Just a common symbol for a covariance.}}$$



Elliptical or Ellipsoidal contours.

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

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

$y = A\underline{x}$

$\underline{x} = A^{-1}y$

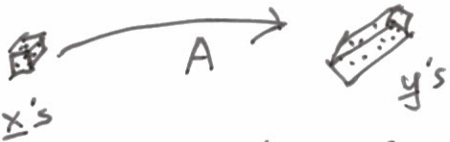
(If A^{-1} exists)

$$\propto e^{-1/2\, y^T \underbrace{A^{-T}A^{-1}}\, y}$$

$$\Sigma = AA^T$$

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

$$\propto e^{-1/2\, y^T \Sigma^{-1} y}$$



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

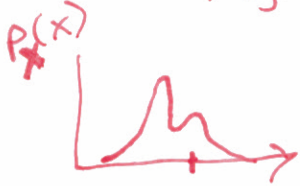
$$|\Sigma| = |AA^T| = |A||A^T| = |A|^2$$

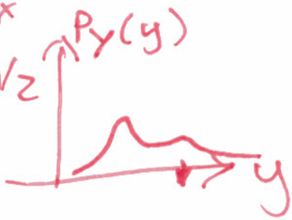
$$= \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-1/2\, y^T \Sigma^{-1} y} = N(y; 0, \Sigma)$$

$$P(y) = P(x(y))$$

$$y = 2x$$

$$x = y/2$$





$$P_Y(y) = |J| \, P_X(x(y))$$

$$\begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_D} \\ \\ \frac{\partial x_D}{\partial y_1} & & \cdots & \frac{\partial x_D}{\partial y_D} \end{vmatrix}$$

$$p_Y(y) \cancel{dy} = p_X(x) \frac{dx}{dy}$$

General Gaussian

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

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

$$= \mathcal{N}(\underline{z}; \underline{\mu}, \Sigma) \leftarrow AA^T$$

<u>Variances are +ve</u>, <u>Covariance usually</u> <u>+ve definite</u>

Iff $\Sigma = AA^T$, Σ is symmetric positive <u>semi</u>-definite

Positive Definite:

$$\underline{z}^T\Sigma\underline{z} > 0 \quad \text{for all } \underline{z} \neq \underline{0}$$

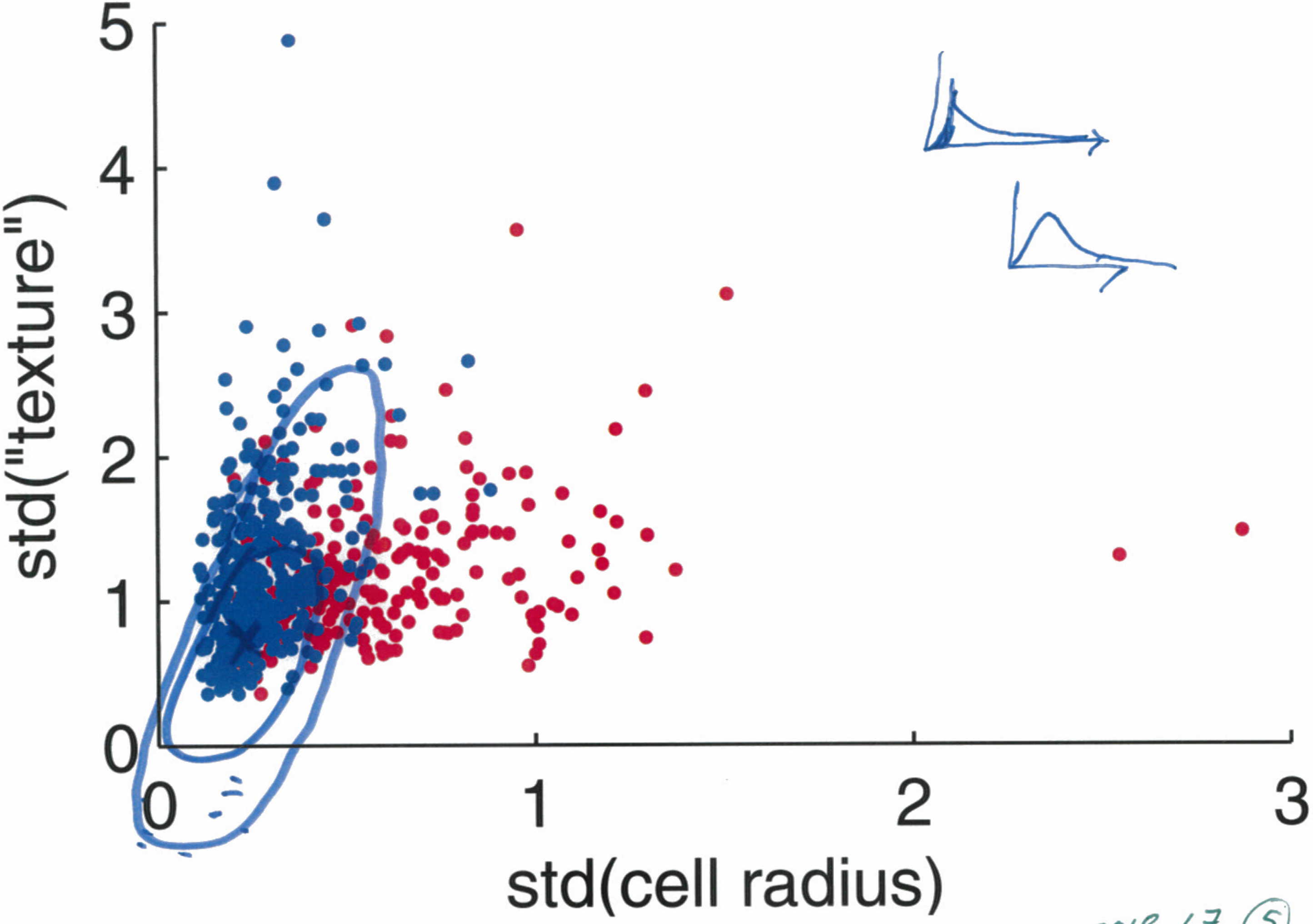
$$\Longleftrightarrow \quad \underline{z}^T\Sigma^{-1}\underline{z} > 0 \quad \text{" " "}$$

[True if A invertible]

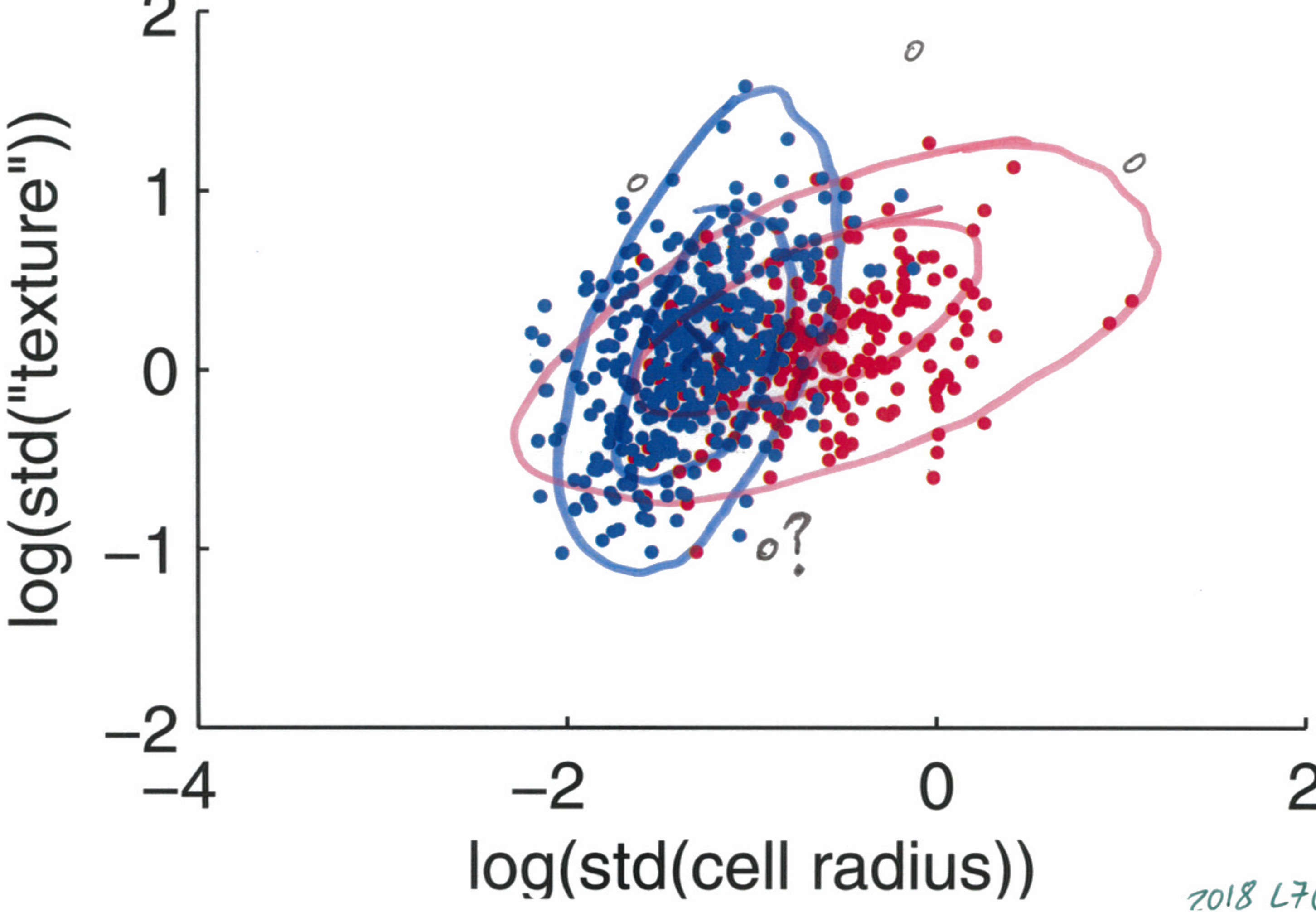
+ve <u>Semi</u>-definite

$$\underline{z}^T\Sigma\underline{z} \geqslant 0 \quad \text{for all } \underline{z}$$

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



2018 L7 5



2018 L7(6)

Bayes Classifier

At training time create model:

Features $p(\underline{x} \mid y=k)$ eg $N(\underline{x}; \underline{\mu}^{(k)}, \Sigma^{(k)})$

(or discrete distribution

or Naive Bayes Classifier

features $\{x_d\}$ are

independent given class)

Labels $p(y=k) = \pi_k$

$(\sum_k \pi_k = 1)$

Fit the model $\pi_k \approx \frac{\#\text{ points in class } k}{N}$

$\underline{\mu}^{(k)}, \Sigma^{(k)}$ set mean and covariance of points in class k.

At test time use Bayes' rule

$$p(y\,|\,\underline{x}) \propto p(y, \underline{x}) \propto p(y)\,p(\underline{x}\,|\,y)$$

$$p(y=k\,|\,\underline{x}) = \frac{\overbrace{p(y=k)}^{\pi_k}\ \overbrace{p(\underline{x}\,|\,y)}^{\mathcal{N}(\underline{x};\,\mu^{(k)},\,\Sigma^{(k)})}}{\sum_{k'} p(y=k')\,p(\underline{x}\,|\,y=k')}$$