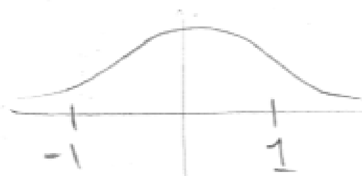
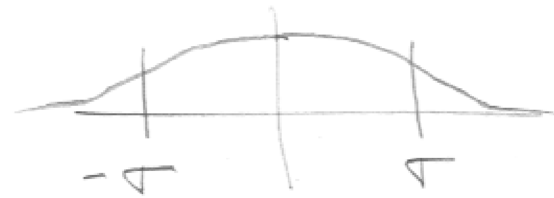
1D GAUSSIAN

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

$N(0,1)$: $G(x) = \frac{1}{\sqrt{2\pi}} e^{-1/2\, x^2}$

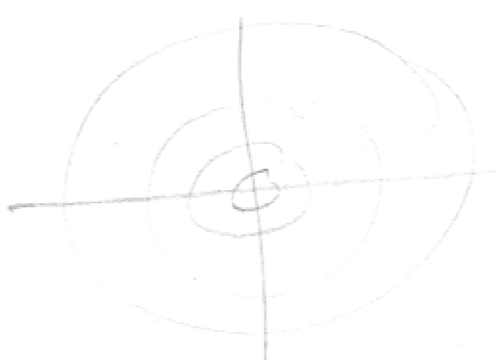


AREA $= \sqrt{2\pi}$



AREA $\times\, \sigma$ $= \sigma\sqrt{2\pi}$

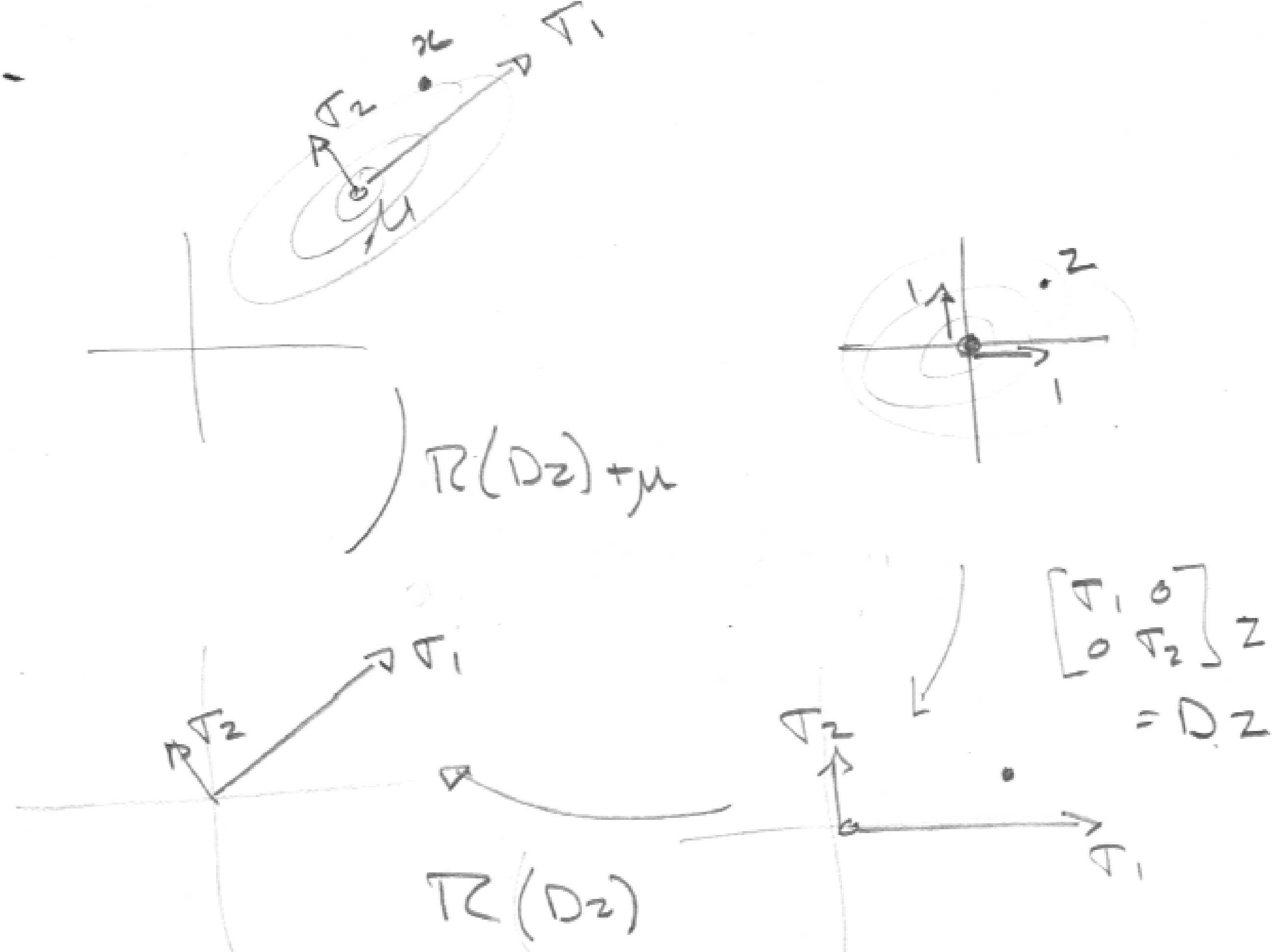
2D GAUSSIAN



$N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1&0\\0&1\end{bmatrix}\right) \approx G\left(\begin{bmatrix}x\\y\end{bmatrix}\right) =$

$$\frac{1}{\sqrt{2\pi}} e^{-1/2\, x^2} \times \frac{1}{\sqrt{2\pi}} e^{-1/2\, y^2}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^2 e^{-1/2\,(x^2+y^2)}$$



$$x = \overbrace{\boxed{RD}}^{A} z + \mu$$

$$z = (RD)^{-1}(x - \mu)$$

$$= A^{-1}(x-\mu)$$

$$\Sigma^{-1} = (AA^T)^{-1} = A^{T^{-1}} A^{-1}$$

$N(\mu, \Sigma)$:

AA^T

$$G(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^2 \frac{1}{\sigma_1 \sigma_2} e^{-1/2 \underbrace{\left|A^{-1}(x-\mu)\right|^2}_{\text{MOVE } x \text{ TO } N_x(0,1) \times N_y(0,1) \text{ DISTRIBUTION}}}$$

$$|u|^2 = u^T u$$

$$\begin{aligned}
|A(x-\mu)|^2 &= (A^{-1}(x-\mu))^T (A^{-1}(x-\mu)) \\
&= (x-\mu)^T A^{-1^T} A^{-1} (x-\mu) \\
&= (x-\mu)^T \Sigma^{-1} (x-\mu)
\end{aligned}$$

$[\text{DET } X^T = \text{DET } X]$

$$\begin{aligned}
\text{DET}(\Sigma) &= \text{DET}(AA^T) \\
&= \text{DET}(A)^2 \\
&= (\text{DET}(R)\ \text{DET}(D))^2 \\
&= 1 \times \text{DET}(D)^2 = \text{DET}\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}^2 \\
&= \sigma_1^2 \sigma_2^2
\end{aligned}$$

$N(\mu, \Sigma)$:

$$G(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^2 \frac{1}{\sqrt{\det \Sigma}} \, e^{-1/2 (x-\mu)^T \Sigma^{-1} (x-\mu)}$$

WORKS IN ANY DIMENSION, k, USING $\left(\frac{1}{\sqrt{2\pi}}\right)^k$

RELATION TO SVD

$$\begin{aligned} A A^T &= (RD)(RD)^T \\ &= R D D^T R^T \\ &= R \begin{bmatrix} \sigma_1^2 & \\ & \sigma_2^2 \end{bmatrix} R^T \end{aligned}$$

$$x = Rz$$

$$= \begin{bmatrix} \wr & \wr \end{bmatrix} z$$

NEW e_1 NEW e_2

PRINCIPAL COMPONENTS

$$A A^T = \begin{bmatrix} \wr & \wr \end{bmatrix} \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \begin{bmatrix} \text{---} \\ \text{---} \end{bmatrix}$$

AA^T IS SYMMETRIC & POSITIVE DEFINITE → $x AA^T x > 0$ FOR $x \neq 0$

$\therefore$ SVD$(AA^T) = U \Sigma U^T$

COLS ARE EIGENVECTORS

DIFFERENT Σ CONTAINING EIGENVALUES

$\Sigma = AA^T$ IS A $\boxed{\text{COVARIANCE MATRIX}}$

RECALL $\text{VAR}(X) = E\left((X - E(X))^2 \right)$

$$= \frac{1}{n} \sum_{x_i} (x_i - \mu)^2$$

$$\text{COV}(X,Y) = E\left((X - E(X))(Y - E(Y)) \right)$$

$$= \frac{1}{n} \sum_{x_i y_i} (x_i - \mu_x)(y_i - \mu_y)$$

$$\Sigma = \begin{bmatrix} \text{cov}(X,X) & \text{cov}(X,Y) \\ \text{cov}(Y,X) & \text{cov}(Y,Y) \end{bmatrix}$$

SYMMETRIC w/ VARIANCES ON DIAGONAL

FROM DATA POINTS

$\boxed{\text{IF CENTROID AT ORIGIN}}$

$$A = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \\ z_1 & z_2 & \cdots & z_n \end{bmatrix}$$

$$AA^T = \begin{bmatrix} | & | & | & \cdots & | \end{bmatrix} \begin{bmatrix} - \\ - \\ - \\ \vdots \\ - \end{bmatrix}$$

$$= \begin{bmatrix} \sum x_i^2 & \sum x_i y_i & \sum x_i z_i \\ \sum y_i x_i & \sum y_i^2 & \sum y_i z_i \\ \sum z_i x_i & \sum z_i y_i & \sum z_i^2 \end{bmatrix}$$

= COVARIANCE MATRIX $\times n$

CAN FIND GAUSSIAN TO FIT A POINT SET WITH $\Sigma = AA^T$

THEN $\mathrm{SVD}(\Sigma) = \underset{\text{AXES}}{R}\ \underset{\text{VARIANCES}}{D}\ R^T$

GAUSSIAN MIXTURE MODELS

EXAMPLES: SIMPSON PAPER
VERBEEK PAPER

ONE GAUSSIAN

$\phi(x, \theta)$

θ DENOTES μ, Σ

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^{k} (\text{DET } \Sigma)^{-1/2} e^{-1/2 (x-\mu)^T \Sigma^{-1} (x-\mu)}$$

GAUSSIAN MIXTURE

k COMPONENTS

$$f_k(x) = \sum_{j=1}^{k} \pi_j \phi(x; \theta_j)$$

WEIGHTS SUM TO 1: $\sum \pi_j = 1$

$\pi_j \geq 0$

π_j = "MIXING WEIGHTS"

$\phi(x; \theta_j)$ = "j^{TH} COMPONENT"

$f_k(x)$ IS A PROBABILITY DENSITY FUNCTION

GIVEN SAMPLE POINTS $\{x_1 \dots x_n\} = X_n$

WE WANT TO MAXIMIZE THE LOG-LIKELIHOOD OF $f_k(x)$

$\mathcal{L}(X_n, f_k)$ = LOG-LIKELIHOOD OF SEEING X_n IN THE k-COMPONENT MIXTURE MODEL f_k

$$P(X_n | f_k) = P(x_1 | f_k) \times P(x_2 | f_k) \times \dots \times P(x_n | f_k)$$

$$\text{LOG}(P(X_n | f_k)) = \sum_{i=1}^{n} \text{LOG}(x_i | f_k)$$

$$= \sum_{i} \text{LOG}\, f_k(x_i)$$

ITERATE THROUGH COMPONENTS

FOR $j = 1$ TO k

OPTIMIZE COMPONENT j FOR X_n

REPEAT

TO OPTIMIZE COMPONENT j

$$P(j \mid x_i) = \frac{\pi_j \, \phi(x_i, \theta_j)}{\sum_{l=1}^{k} \pi_l \, \phi(x_i, \theta_l)} \qquad \sum_{l=1}^{k} \pi_l \, \phi(x_i, \theta_l) = f_k(x_i)$$

PROB THAT x_i BELONGS TO j

$$\pi_j = \frac{1}{n} \sum_{i=1}^{n} P(j \mid x_i)$$

AVG OF PROBS THAT x_i BELONGS TO j
I.E. PROB THAT <u>ALL</u> X_n BELONG TO j

THIS WILL KEEP $\sum \pi_j = 1$
AFTER $j = 1 \ldots k$

$$\mu_j' = \frac{\frac{1}{n}\sum_{i=1}^{n} P(j|x_i)\, x_i}{\frac{1}{n}\sum P(j|x_i)}$$

$$\frac{1}{n}\sum P(j|x_i) = \pi_j$$

PROBABILITY-WEIGHTED MEAN OF j

$$C_j = \frac{1}{n\pi_j}\sum_{i=1}^{n} P(j|x)\,\underbrace{(x_i-\mu_j)(x_i-\mu_j)^T}$$

$$\begin{bmatrix} a \\ b \end{bmatrix} \times \begin{bmatrix} a & b \end{bmatrix}$$

$$= \begin{bmatrix} a^2 & ab \\ ba & b^2 \end{bmatrix}$$

= COVARIANCE OF x_i WITH ITSELF

RECALL $\begin{bmatrix} a & a & \cdots & a \\ b & b & \cdots & b \end{bmatrix}$ $\begin{bmatrix} a\,b \\ a\,b \\ \vdots \\ a\,b \end{bmatrix}$

$$\sum$$

COMPUTES COVARIANCE MATRIX!

GREEDY VARIANT

FIND f_1

GIVEN f_k, ADD ONE COMPONENT

$$f_{k+1} = (1-\alpha) f_k + \alpha \phi(x, \theta_{k+1})$$

USE EM TO OPTIMIZE α & θ_{k+1},

BUT NOT $\pi_1 \ldots \pi_k$ OR $\theta_1 \ldots \theta_k$

FOR OPTIMAL α^* AND θ^*, SET

SET EACH $\pi_j \leftarrow (1-\alpha^*) \pi_j$ $\quad j = 1 \ldots k$

SET $\pi_{k+1} \leftarrow \alpha^*$

SET $\theta_{k+1} \leftarrow \theta^*$