L9: Linear Gaussian models

So far we've focused on discrete-valued models, e.g. binary.
What about continuous variables and models?
Much of the same formalism applies.

Maximum likelihood for a Gaussian.

$$p(\underline{x}|\mu,\sigma^2) = \prod_{n=1}^{N} \mathcal{N}(x_n|\mu,\sigma^2)$$

How do we find best μ?

$$\ln p(\underline{x}|\mu,\sigma^2) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \text{const.}$$

$$\frac{\partial \sim}{\partial \mu} = -\frac{1}{2\sigma^2}\sum_{n=1}^{N} 2(x_n-\mu)(-1) = 0$$

$$= \frac{1}{\sigma^2}\sum_{n=1}^{N} x_n - \underbrace{\sum_{n=1}^{N}\mu}_{=N\mu} = 0$$

$$\Rightarrow \quad \mu = \frac{1}{N}\sum_{n=1}^{N} x_n$$

② L9: LGMs

$p(\underline{w}|\alpha)$

$$\prod_j \left(\frac{\alpha}{2\pi}\right)^{1/2} \exp\left[-\frac{\alpha}{2} w_j^2\right]$$

$$= \left(\frac{\alpha}{2\pi}\right)^{\frac{(M+1)}{2}} \exp\left[-\frac{\alpha}{2} w^T w\right]$$

$$\sum \left(-\frac{1}{2} w_j^2\right) = -\frac{1}{2} \underline{w}^T w$$

---

predictive distribution

$$p(t|x, \underline{x}, \underline{t}) = ?$$

$$p(w|\underline{x}, \underline{t}) = \frac{p(\underline{t}|\underline{x}, \underline{w}, \beta)\, p(\underline{w}|\alpha)}{p(\underline{t}|\underline{x})}$$

$$= \int p(\underline{t}|\underline{x}, \underline{w}, \beta)\, p(\underline{w}|\alpha)\, d\underline{w}$$

→ Gaussian × Gaussian = Gaussian
so we can compute this.

$$p(t|x, \underline{w})\, p(\underline{w}|\underline{x}, \underline{t})$$

$$= p(t, \underline{w}|x, \underline{x}, \underline{t})$$

$$p(t|x, \underline{x}, \underline{t}) = \int d\underline{w}\; p(t, \underline{w}|x, \underline{x}, \underline{t})$$

Gaussian, so can be performed analytically (B3.3)

Multivariate Gaussians:

Recall univariate:

$$\mathcal{N}(x \mid \mu, \sigma^2) \quad \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$$

Now multivariate:

$$\mathcal{N}(\underline{x} \mid \underline{\mu}, \underline{\underline{\Sigma}}) = \underbrace{\frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}}}_{\text{normalizing const.}} \exp\left[-\frac{1}{2}\overbrace{(\underline{x}-\underline{\mu})^T \underline{\underline{\Sigma}}^{-1} (\underline{x}-\underline{\mu})}^{\Delta^2}\right]$$

Δ = Manhalanobis distance from $\underline{x}$ to $\underline{\mu}$

= Euclidean dist when $\underline{\underline{\Sigma}} = \underline{\underline{I}}$

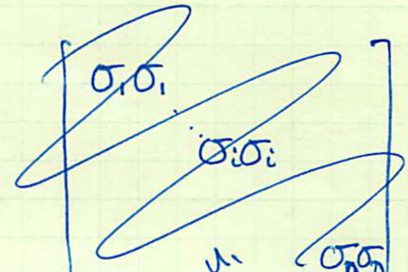
i.e. $\|\underline{x} - \underline{\mu}\|$

$\underline{\mu}$ = D-dim. mean vector

$\underline{\underline{\Sigma}}$ = D×D covariance matrix

$|\Sigma|$ = determinant of Σ

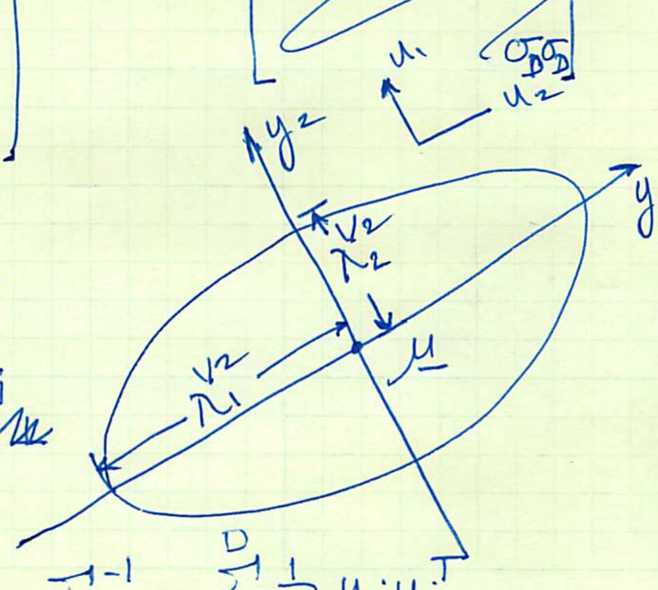
How do you understand $\underline{\underline{\Sigma}}$?



each entry in matrix is the covariance $\underline{\underline{\Sigma}}_{ij}$ between x_i & x_j

$$\begin{bmatrix} \ddots & & \\ & \sigma_i\sigma_j & \\ \cdot & & \cdot \end{bmatrix}$$

Look at density contours.



$$\underline{\underline{\Sigma}}\, \underline{u}_i = \lambda_i \underline{u}_i$$

λ_i ← evals, $\underline{u}_i$ ← evecs

$\underline{u}_i$ can be choosen so $u_i^T u_j = I_{ij}$

Can write Σ in terms of evecs ~~form~~

$$\underline{\underline{\Sigma}} = \sum_{i=1}^{D} \lambda_i \underline{u}_i u_i^T \Rightarrow \Sigma^{-1} = \sum_{i=1}^{D} \frac{1}{\lambda_i} u_i u_i^T$$

④ L9: LGM:

Can rewrite $\Delta^2 = (\underline{x}-\underline{\mu})^T \underline{\underline{\Sigma}}^{-1} (\underline{x}-\underline{\mu})$ as

$$\Delta^2 = \sum_{i=1}^{D} \frac{y_i^2}{\lambda_i} \qquad y_i = \underbrace{u_i^T(\underline{x}-\underline{\mu})}$$

i.e. projection of $\underline{x}-\underline{\mu}$ onto evecs

y_i is a new coordinate system which is equivalent to having D independent vars.

In vec form

$$\underline{y} = \underline{\underline{U}}(\underline{x}-\underline{\mu})$$

Also note $|\Sigma| = \prod_{i=1}^{D} \lambda_i$

so $|\Sigma|^{1/2} = \prod_{i=1}^{D} \lambda_i^{1/2}$

U is an orthogonal matrix so

$\underline{\underline{U}}\,\underline{\underline{U}}^T = \underline{\underline{I}}$

and $\underline{\underline{U}}^T\underline{\underline{U}} = \underline{\underline{I}}$

So we can rewrite the multivariate as

$$p(\underline{y}) = \prod_{j=1}^{D} \frac{1}{(2\pi\lambda_j)^{1/2}} \exp\left[-\frac{y_j^2}{2\lambda_j}\right]$$

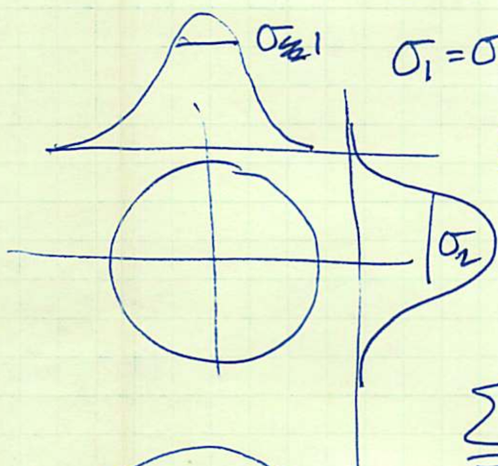
TOPS 35500

① 491 L10 – Gaussian Models

multivariate Gaussians:

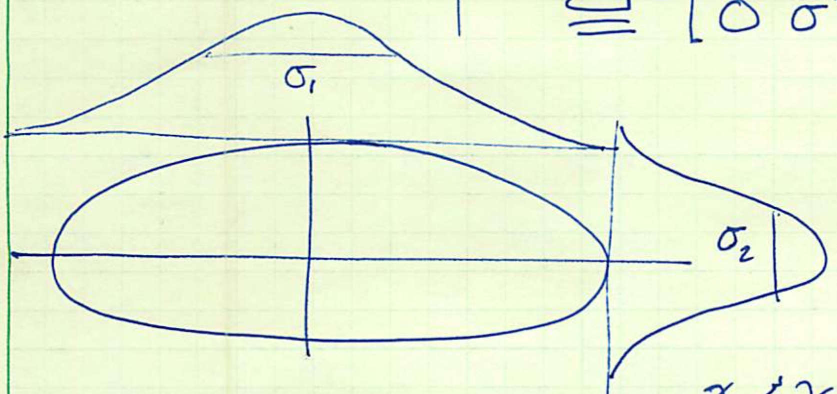
$$\mathcal{N}(x \mid \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$$

$$\mathcal{N}(\underline{x} \mid \underline{\mu}, \underline{\underline{\Sigma}}) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left[\underbrace{-\frac{1}{2}(\underline{x}-\underline{\mu})^T \underline{\underline{\Sigma}}^{-1}(\underline{x}-\underline{\mu})}_{=\Delta^2}\right]$$



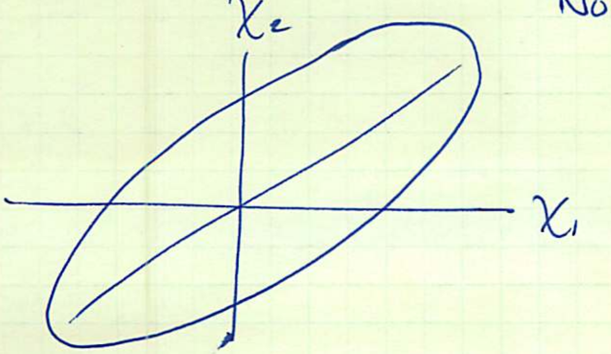
$\sigma_1 = \sigma_2 = \sigma$

$$\underline{\underline{\Sigma}} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$$



$$\underline{\underline{\Sigma}} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

x_1 & x_2 are now correlated.



$$\underline{\underline{\Sigma}} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

$$\sigma_{12} = \sigma_{21}$$

$\underline{\underline{\Sigma}}$ is symmetric $\Rightarrow \underline{\underline{\Sigma}} = \underline{\underline{\Sigma}}^T$

② 491 L10 - Gaussian Models

$$\Delta^2 = (\underline{x}-\underline{\mu})^T \underline{\underline{\Sigma}}^{-1} (\underline{x}-\underline{\mu})$$

$\Delta = \sqrt{\Delta^2}$ = Manhalanobis dist

Δ = Euclidean dist when $\underline{\underline{\Sigma}} = \underline{\underline{I}}$

Eigenvecs:

$$\underline{\underline{A}}\,\underline{u}_i = \underset{\text{eigen value}}{\lambda_i}\,\underset{\text{eigenvector}}{\underline{u}_i}$$

$$\begin{bmatrix} & & \\ & \underline{\underline{A}} & \\ & & \end{bmatrix}\begin{bmatrix} \\ \underline{u}_i \\ \\ \end{bmatrix} = \lambda_i \begin{bmatrix} \\ \underline{u}_i \\ \\ \end{bmatrix}$$

same as multiplying $\underline{\underline{A}}$ by scalar, λ_i = eval.

$\underline{u}_i^T \underline{u}_j = 0$ unless $i=j$ then $=1$

$$\underline{u}_i^T \underline{u}_j = I_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

$$\underline{\underline{\Sigma}} = \sum_{i=1}^{D} \lambda_i \underline{u}_i \underline{u}_i^T = \lambda_1 \begin{bmatrix} \\ \\ \\ \end{bmatrix}\begin{bmatrix} & & \end{bmatrix} = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}$$

sum of rank 1 matrices

$$\underline{\underline{\Sigma}}^{-1} = \sum_{i=1}^{D} \frac{1}{\lambda_i} \underline{u}_i \underline{u}_i^T$$

$$\Rightarrow \Delta^2 = (\underline{x}-\underline{\mu})^T \left[\sum_{i=1}^{D} \frac{1}{\lambda_i} \underline{u}_i \underline{u}_i^T \right] (\underline{x}-\underline{\mu})$$

$$= \sum_{i=1}^{D} \frac{1}{\lambda_i} \underbrace{(\underline{x}-\underline{\mu})^T \underline{u}_i}_{=y_i} \underbrace{\underline{u}_i^T (\underline{x}-\underline{\mu})}_{=y_i} = \sum_{i=1}^{D} \frac{y_i^2}{\lambda_i}$$

Now: $y_i \sim \mathcal{N}(\mu_i, \lambda_i) \Rightarrow p(\underline{y} \mid \underline{\mu}, \underline{\lambda}) = \prod_{i=1}^{D} p(y_i \mid \mu_i, \lambda_i)$

Transforming $\underline{x}$ into the eigenspace converts the $\underline{\underline{\Sigma}}$ to a set of univariate Gaussians.

$$\Delta^2 = \sum_{i=1}^{D} \frac{y_i^2}{\lambda_i} \qquad |\underline{\underline{\Sigma}}|^{1/2} = \left(\prod_i \lambda_i\right)^{1/2} = \prod_i \lambda_i^{1/2}$$

$$\Rightarrow p(\underline{x} \mid \underline{\mu}, \underline{\underline{\Sigma}}) =$$

$$\frac{1}{(2\pi)^{D/2} |\underline{\underline{\Sigma}}|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu})^T \underline{\underline{\Sigma}}^{-1} (\underline{x}-\underline{\mu})\right]$$

$$= \frac{1}{2\pi^{D/2} \prod_i \lambda^{1/2}} \exp\left[-\frac{1}{2} \sum_{i=1}^{D} \frac{y_i^2}{\lambda_i}\right]$$

$$p(y_i \mid \mu_i = 0, \sigma^2 = \lambda_i) = \frac{1}{(2\pi\lambda_i)^{1/2}} \exp\left[-\frac{1}{2}\frac{y_i^2}{\lambda_i}\right]$$

$$p(\underline{y} \mid \underset{\substack{\downarrow \\ \text{was } \underline{\mu}}}{\underline{\mu}} = 0, \underline{\lambda}) = \prod_{i=1}^{D} p(y_i \mid \mu_i = 0, \sigma^2 = \lambda_i)$$

$$= \prod_{i=1}^{D} \frac{1}{(2\pi\lambda_i)^{1/2}} \exp\left[-\frac{1}{2}\frac{y_i^2}{\lambda_i}\right]$$

$$= \frac{1}{(2\pi)^{D/2} \prod_i \lambda_i^{1/2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{D} \frac{y_i^2}{\lambda_i}\right]$$

Each $y_i = u_i^T(\underline{x} - \underline{\mu})$

In matrix form $\underline{y} = \underline{\underline{U}}(\underline{x} - \underline{\mu})$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_D \end{bmatrix} = \begin{bmatrix} — \underline{u}_1^T — \\ \vdots \\ — \underline{u}_D^T — \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ \vdots \\ x_D - \mu_D \end{bmatrix}$$

Some properties of multivariate Gaussians

Let $\underline{x} = \begin{bmatrix} \underline{x}_a \\ \underline{x}_b \end{bmatrix}$ partition $\underline{x}$ into two subsets

$$\underline{\mu} = \begin{bmatrix} \underline{\mu}_a \\ \underline{\mu}_b \end{bmatrix} \qquad \underline{\underline{\Sigma}} = \begin{bmatrix} \underline{\underline{\Sigma}}_{aa} & \underline{\underline{\Sigma}}_{ab} \\ \underline{\underline{\Sigma}}_{ba} & \underline{\underline{\Sigma}}_{bb} \end{bmatrix}$$

$$\underline{\underline{\Sigma}}^{-1} = \underline{\underline{\Lambda}} = \begin{bmatrix} \underline{\underline{\Lambda}}_{aa} & \underline{\underline{\Lambda}}_{ab} \\ \underline{\underline{\Lambda}}_{ba} & \underline{\underline{\Lambda}}_{bb} \end{bmatrix}$$

What is $p(\underline{x}_a | \underline{x}_b)$? Turns out it's also Gaussian.

$$p(\underline{x}_a | \underline{x}_b) = \mathcal{N}(\underline{\mu}_{a|b}, \underline{\underline{\Sigma}}_{a|b})$$

$$\underline{\mu}_{a|b} = \underline{\mu}_a - \underline{\underline{\Lambda}}_{aa}^{-1} \underline{\underline{\Lambda}}_{ab} (\underline{x}_b - \underline{\mu}_b)$$

$$\underline{\underline{\Sigma}}_{a|b} = \underline{\underline{\Lambda}}_{aa}^{-1} \quad (\neq \underline{\underline{\Sigma}}_{aa}^{-1})$$

(see slides)

TOPS 35500