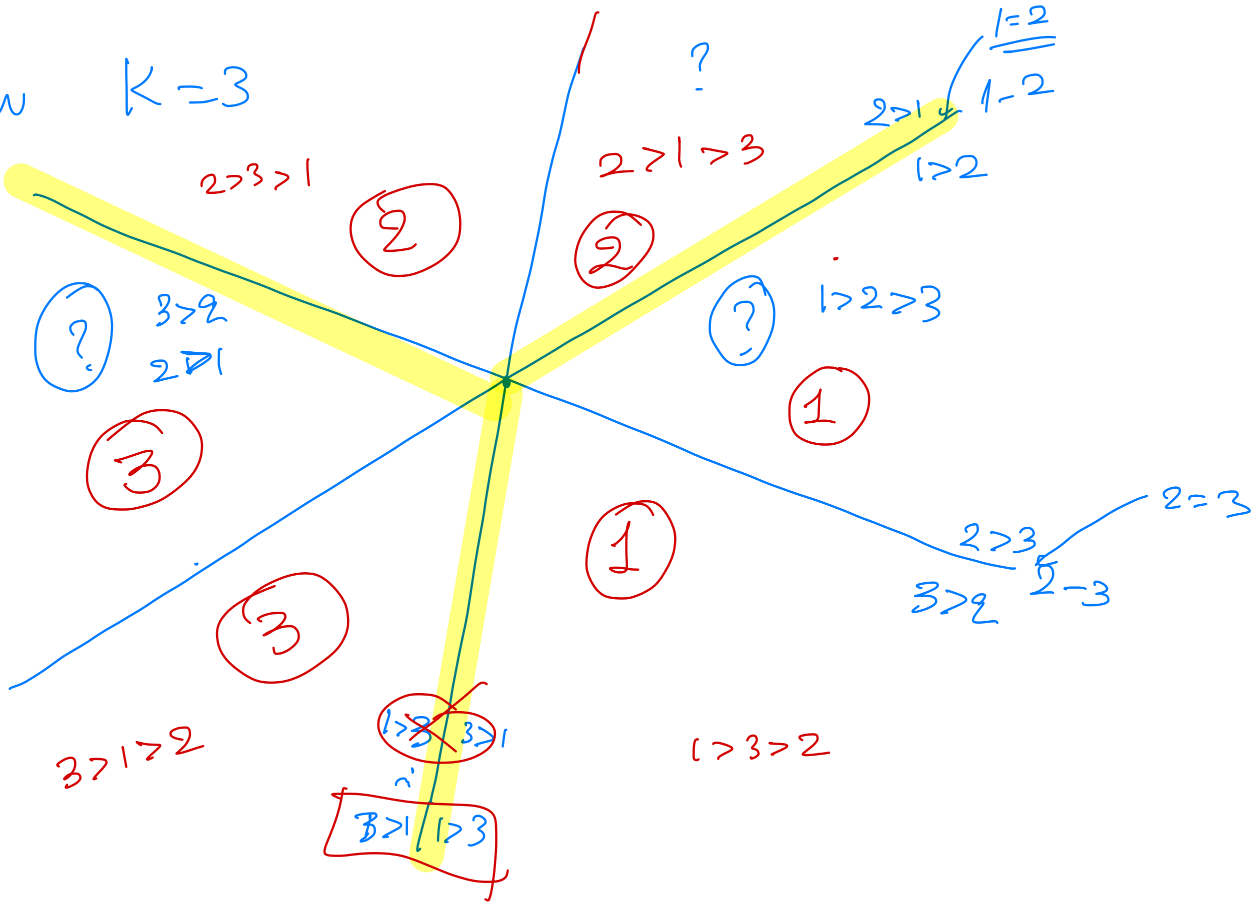
20-10-2021

Γραμμικά Σύνορα

$\hat{y}_k > \hat{y}_\ell$ $\hat{y}_k$ = γραμμικά ως προς x

π.χ. αν $K=3$



Έχουμε δει classification μέσω regression

Επιστρέφουμε στο Bayesian model

$\pi = (\pi_1, \dots, \pi_K)$: prior probabilities $P(G=k)$

Αν ξέρουμε την κατανομή $\underline{(G, X)}$

Bayes classifier

$$y(x) \in \underset{k\in K}{\arg\max}\ \underline{P(G=k \mid X=x)}$$

$$\boxed{\begin{array}{lll} \underline{\text{Κατανομή}} & X \mid G=k & : \text{pdf } f_k(x) \\ \text{Prior} & \pi_k = P(G=k) & \end{array}} \Rightarrow$$

$$P(G=k \mid X=x) = \frac{P(G=k)\, f_k(x)}{\sum_\ell P(G=\ell)\, f_\ell(x)} = C(x) \underbrace{\pi_k f_k(x)}_{L(k,x)}$$

$$\boxed{P(G=k \mid x) > P(G=\ell \mid x)} \Rightarrow$$

← πότε καταλήγει σε γραμμικά όρια?

$$\pi_k f_k(x) > \pi_\ell f_\ell(x) \Rightarrow \frac{\pi_k f_k(x)}{\pi_\ell f(x)} > 1$$ λόγος πιθανοφάνειας

$$\Rightarrow \log\frac{\pi_k}{\pi_\ell} + \boxed{\log\frac{f_k(x)}{f_\ell(x)}} > 0$$ σύνορο μεταξύ k κ ℓ

① Discriminant Analysis (Linear)

υποθέτοντας κανονική κατανομή για τα $X|G=k$

② Logistic Regression

απευθείας τα $P(G=k|X=x)$

με κατάλληλο τρόπο

Linear Discriminant Analysis

Διαχωριστική Ανάλυση

$X \in \mathbb{R}^p$

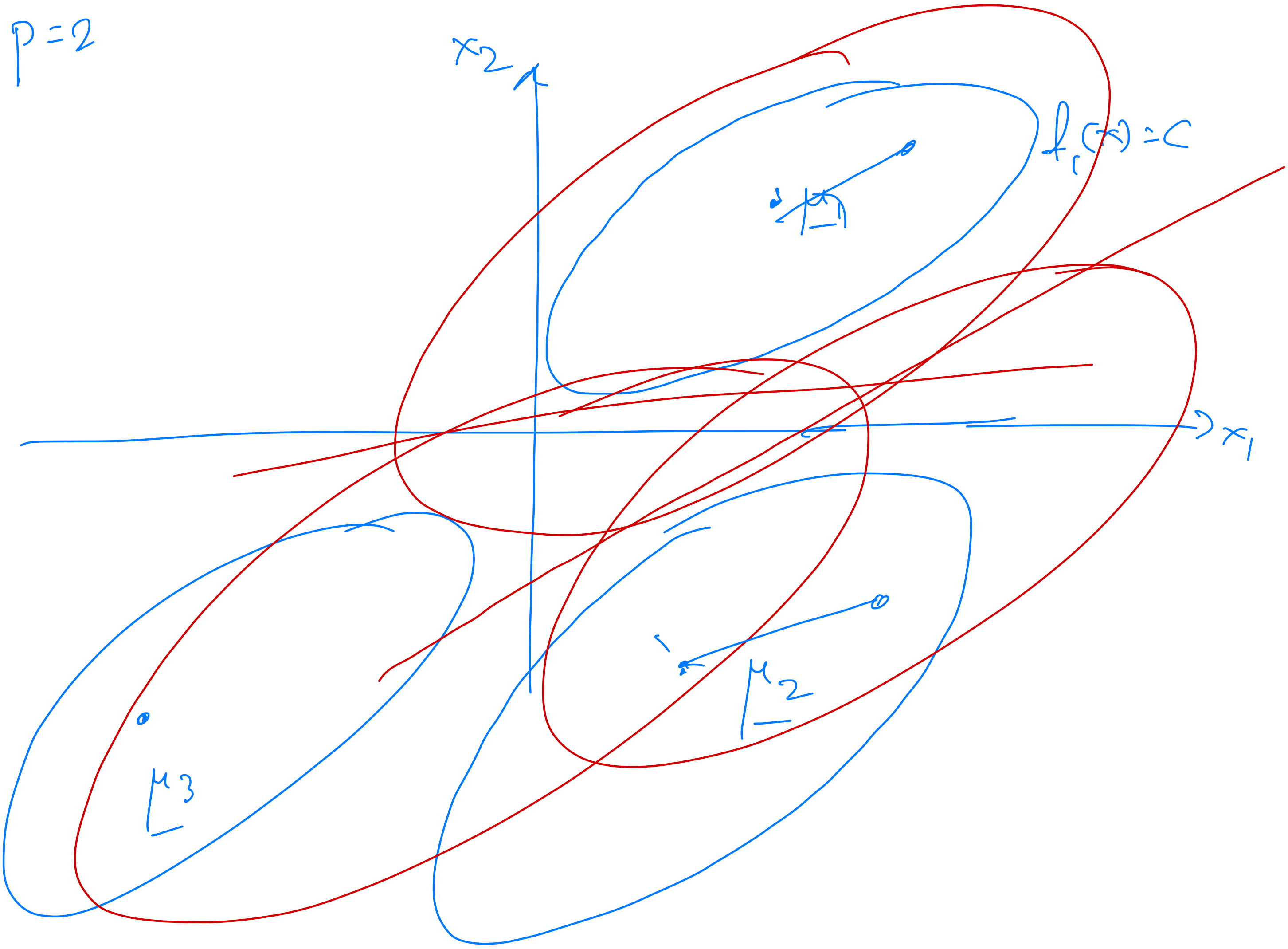
centroids → $\mu_k \in \mathbb{R}^p$

Υπόθεση $X|G=k \sim \mathcal{N}(\mu_k, \Sigma_k)$ Σ_k $p\times p$

i) Βασική υπόθεση $\Sigma_k = \Sigma$ (οδηγεί σε γραμμικά σύνορα)

$$\boxed{f_k(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T\Sigma^{-1}(x-\mu_k)}}$$

$p=2$



Πώς προκύπτουν τα σύνορα?

Πρέπει να υπολογίσουμε τα

$$\underline{\log \frac{\pi_k}{\pi_\ell} + \log \frac{f_k(x)}{f_\ell(x)} = 0}$$ σύνορο $k-\ell$

$$\log f_k(x) = \log\left(\frac{1}{(2\pi)^{1/2}|\Sigma|^{1/2}}\right) - \frac{1}{2}(x-\mu_k^T)\Sigma^{-1}(x-\mu_k)$$

$$\boxed{\log P(G=k|X=x)} = \boxed{\log \pi_k + (\quad)} - \frac{1}{2}(x-\mu_k^T)\Sigma^{-1}(x-\mu_k)$$

$$\log\frac{f_k(x)}{f_\ell(x)} = \log\frac{\pi_k}{\pi_\ell} - \frac{1}{2}\left[(x-\mu_k)^T\Sigma^{-1}(x-\mu_k) - (x-\mu_\ell)^T\Sigma^{-1}(x-\mu_\ell)\right]$$

$$\cancel{x^T\Sigma^{-1}x} + \cdots \quad - \cancel{x^T\Sigma^{-1}x} + \cdots$$

$$= \boxed{\log\frac{\pi_k}{\pi_\ell} + x^T\Sigma^{-1}(\mu_k-\mu_\ell) - \frac{1}{2}(\mu_k+\mu_\ell)^T\Sigma^{-1}(\mu_k-\mu_\ell)}$$

σύνορο $k-\ell$

Σ: συμμετρικό $\qquad x^T\Sigma^{-1}\mu_k = (\mu^k)^T\Sigma^{-1}x$

Δ.Ο. $\quad \log\dfrac{P(G=k|x)}{P(G=\ell|x)} > 0 \iff \delta_k(x) > \delta_\ell(x)$

όπου $\quad \boxed{\delta_k(x) = x^T\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + \log\pi_k}$

διακρίνουσα συνάρτηση της κατηγορίας K (discriminant function)

Given $X=x$ πρόβλεψη $\qquad \boxed{\hat{G}(x) \in \arg\max_k \delta_k(x)}$ πεπερασμένο σύνολο

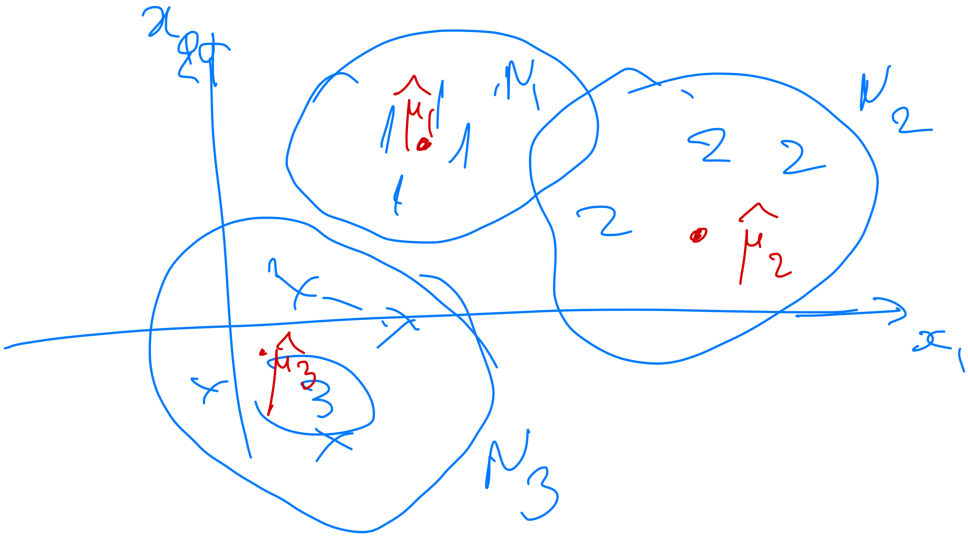
Bayes classifier

Τι γίνεται αν δεν γνωρίζουμε τα π_k, μ_k, Σ ?

Πρέπει να εκτιμηθούν

Έστω training set μεγέθους N

N_k = αρ. παρατηρ. με $G=k$, $N_1+\cdots+N_K=N$



$$\hat{\mu}_k = \frac{1}{N_k}\sum_{i:G=k} x_i$$

$$\hat{\pi}_k = \frac{N_k}{N}$$

$$\hat{\Sigma} = \frac{\sum_{k=1}^{K}\sum_{i:G=k}(x_i-\hat{\mu}_k)(x_i-\hat{\mu}_k)^T}{N-K}$$

Αν $K=1$

$$\hat{\Sigma} = \frac{\sum (x_i-\hat{\mu})(x_i-\hat{\mu})^T}{N-1}$$

$$\Longrightarrow \boxed{\delta_k(x) \quad k=1,\dots,K}$$

(ii) Αν υποθέσουμε γενικά Σ_k $k=1,\dots,K$.

$$\log \frac{f_k(x)}{f_\ell(x)} = x^T \Sigma_k^{-1} x - x^T \Sigma_\ell^{-1} x \dots - - -$$

τετραγωνικά σύνορα.

Quadratic Discriminant Analysis.

Αν υποθέσουμε $\Sigma_k = \Sigma \quad \forall k$

Εκτίμηση

$\pi_1, \dots, \pi_K \in \mathbb{R}$ K

$\mu_1, \dots, \mu_K \in \mathbb{R}^p$ $K \times p$

$\Sigma \in \mathbb{R}^{p \times p}$ p^2

Αν Σ_k διαφ $K \cdot p^2$

Υπολογισμοί

Όροι $\quad (x-\hat{\mu}_k)^T \hat{\Sigma}_k^{-1} (x-\mu_k)$

Διαγωνιοποίηση $\quad \hat{\Sigma}_k : U_k D_k U_k^T \quad$ U_k $p\times p$ ορθοκανονικοί πίνακες.

$D_k = \mathrm{diag}(d_1, \ldots, d_p)$ ιδιοτιμές

$$\hat{\Sigma}_k^{-1} = U_k D_k^{-1} U_k^T$$

$$(x-\hat{\mu}_k)^T \Sigma_k^{-1} (x-\mu_k) = \left(U_k^T (x-\hat{\mu}_k)\right)^T D_k^{-1} \left(U_k^T (x-\hat{\mu}_k)\right)$$

· Μετασχηματισμός $\quad \boxed{X^* = D^{-1/2} U^T X}$

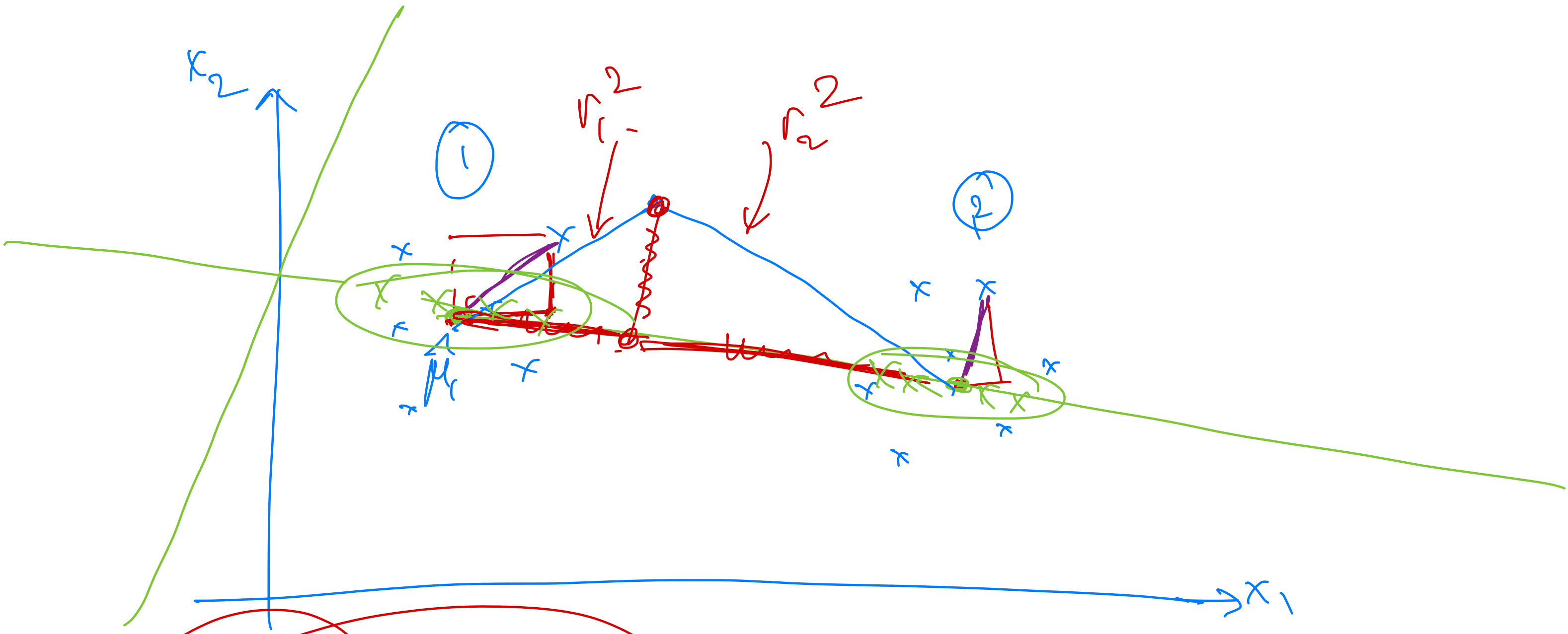
↓

σφαιρικοποιημένα

$$\mathrm{Var}(X^*) = I$$

Παρατήρηση

Έστω $K=2$, $p=2$



$(x-\hat{\mu}_1)^T \cdot (x-\hat{\mu}_1)$ $\quad$ $(x-\hat{\mu}_2)^T(x-\hat{\mu}_2)$

Εφαρμογή όταν $p >> K$.