Gaussian Classifiers

- Let's imagine the Bayesian Decision Rule (BDR) for the case of two multivariate Gaussian case:

$$\boxed{P_{X|Y}(x\,|\,i) = \frac{1}{\sqrt{(2\pi)^d\,|\Sigma_i|}} \exp\left\{-\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)\right\}}$$

  - the BDR

$$\boxed{i^*(x) = \arg\max_i \left[\log P_{X|Y}(x\,|\,i) + \log P_Y(i)\right]}$$

  - becomes

$$\boxed{\begin{aligned} i^*(x) = \arg\max_i \Bigg[& -\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i) \\ & -\frac{1}{2}\log(2\pi)^d\left|\Sigma_i\right| + \log P_Y(i) \Bigg] \end{aligned}}$$

Gaussian Classifiers

- this can be written as

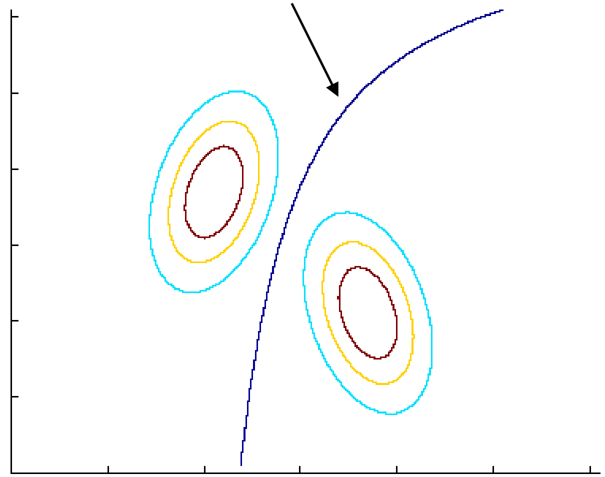
$$\boxed{i^*(x) = \arg\min_i \left[d_i(x,\mu_i) + \alpha_i \right]}$$

with

$$\boxed{d_i(x,y) = (x-y)^T \Sigma_i^{-1} (x-y)}$$

$$\boxed{\alpha_i = \log(2\pi)^d \left| \Sigma_i \right| - 2 \log P_Y(i)}$$

discriminant:
$P_{Y|X}(1|\mathbf{x}) = 0.5$

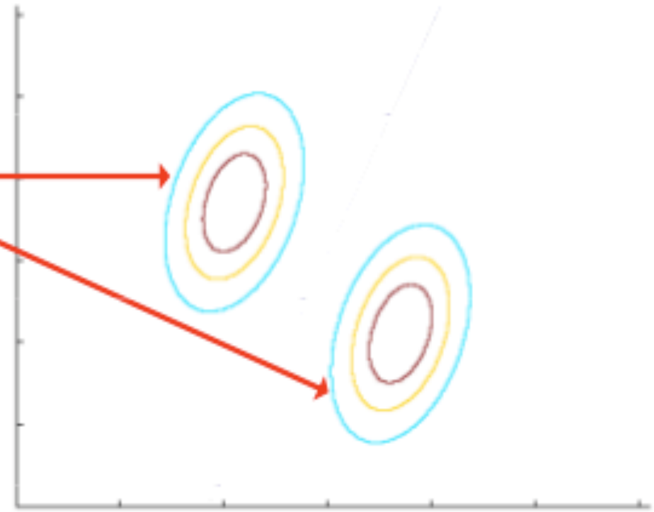


- the optimal rule is to assign x to the closest class
- closest is measured with the Mahalanobis distance $d_i(x,y)$
- to which the α constant is added to account for the class prior

Gaussian Classifiers

- first special case of interest:
  - all classes have the same covariance,

$$\boxed{\Sigma_i = \Sigma, \quad \forall i}$$



- the BDR becomes

$$\boxed{i^*(x) = \arg\min_i \left[d(x,\mu_i) + \alpha_i \right]}$$

- with

$$\boxed{d(x,y) = (x-y)^T \Sigma^{-1} (x-y)}$$

same metric for all classes

$$\boxed{\alpha_i = \log(\cancel{2\pi})^d |\Sigma| - 2\log P_Y(i)}$$

constant, not function of i, can be dropped

Gaussian Classifiers

- In detail:

$$\begin{aligned}
i^*(x) &= \arg\min_i \left[(x-\mu_i)^T \Sigma^{-1}(x-\mu_i) - 2\log P_Y(i)\right] \\
&= \arg\min_i \left[x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu_i - \mu_i^T\Sigma^{-1}x + \mu_i^T\Sigma^{-1}\mu_i - 2\log P_Y(i)\right] \\
&= \arg\min_i \left[x^T\Sigma^{-1}x - 2\mu_i^T\Sigma^{-1}x + \mu_i^T\Sigma^{-1}\mu_i - 2\log P_Y(i)\right] \\
&= \arg\max_i \left[\underbrace{\mu_i^T\Sigma^{-1}}_{w_i^T}x - \underbrace{\frac{1}{2}\mu_i^T\Sigma^{-1}\mu_i + 2\log P_Y(i)}_{w_{i0}}\right]
\end{aligned}$$

Gaussian Classifiers

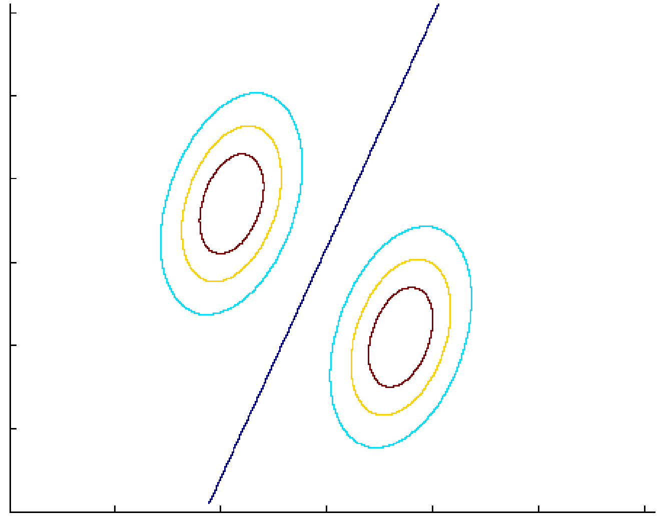
- in summary,

$$\boxed{i^*(x) = \arg\max_i g_i(x)}$$

- with

$$\boxed{\begin{aligned} g_i(x) &= w_i^T x + w_{i0} \\ w_i &= \Sigma^{-1}\mu_i \\ w_{i0} &= -\frac{1}{2}\mu_i^T \Sigma^{-1}\mu_i + \log P_Y(i) \end{aligned}}$$

discriminant:
$P_{Y|X}(1|\mathbf{x}) = 0.5$



- the BDR is a linear function or a linear discriminant

Gaussian Classifiers

- Group Homework 0
  - Find the Geometric equation for the hyperplane separating the two classes for the linear discriminant of the Gaussian classifier.
  - Hint: This is the set such that

$$\boxed{g_i(x) = g_j(x)}$$