TEORÍA DE DECISIÓN BAYESIANA

Estas transparencias contienen material adaptado del libro Duda

DECISIÓN BAYESIANA

- Enfoque estadístico fundamental en clasificación de patrones

- Idea: Estudiar probabilidades de tomar decisiones incorrectas para cuantificar los costos y compromisos de esas decisiones y diseñar las estrategias de menor costo

METODOLOGÍA

1. Supuestas conocidas todas las probabilidades en juego estudiaremos como establecer las reglas de decisión.
2. Posteriormente analizaremos como proceder cuando no se conocen las probabilidades completamente.

Ejemplo: Clasificación de brotes y de hierbas parásitas en cultivos, mediante la captura de imágenes multiespectrales (4 bandas), con el objetivo de realizar una fumigación específica.

EJEMPLO

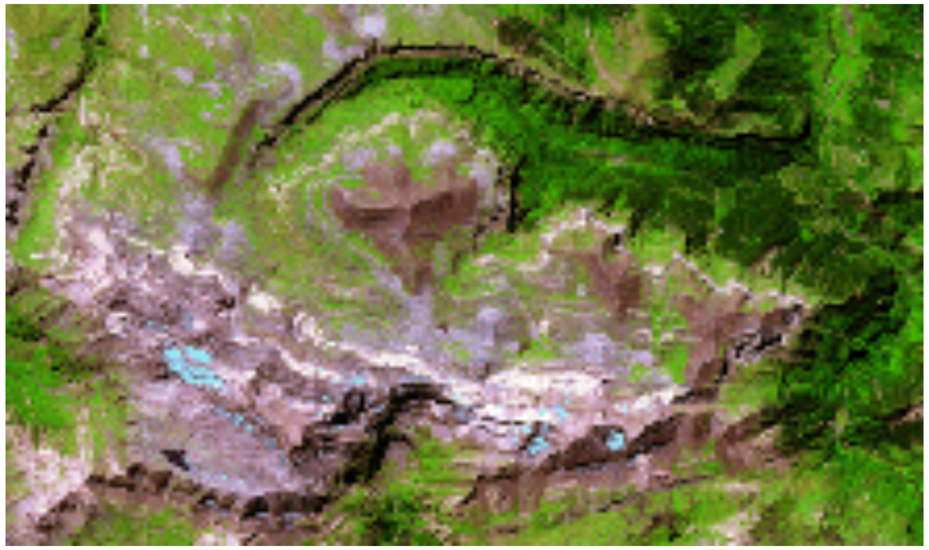
- Clasificación de brotes y de hierbas parásitas en cultivos, mediante la captura de imágenes multiespectrales (4 bandas), con el objetivo de realizar una fumigación específica.
- Pre-procesamiento: discriminación suelo –vegetación
- Problema de clasificación de 2 clases: cada pixel de vegetación pertenece a:

  w_1 – *brote*

  w_2 – *parásito*

  $C=\{ w_1, w_2\}$

  $\Omega \in C$, *V. A*



PRIORS

- $P(w_1)$, $P(w_2)$ probabilidades a priori, pixel brote o parásito. Reflejan conocimiento previo de cuan probable es que un pixel corresponda a brote o parásito antes de inspeccionar imagen.
- Supondremos $P(w_1)+P(w_2)=1$, todo pixel detectado como vegetación es brote o parásito.

REGLA DE DECISIÓN

- Supongamos que somos forzados a tomar una decisión y que todos los costos de decisiones incorrectas son iguales.
- Si la única información a la que podemos acceder son las probabilidades a priori, la regla de decisión razonable es:

$$\boxed{\text{Decido: } \quad w_1 \quad \text{si} \quad P(w_1) > P(w_2)\ ,\ w_2 \ \text{en otro caso}}$$

- Si $P(w_1) >> P(w_2)$ al decidir w_1 casi siempre estamos en lo cierto.
- Si $P(w_1) \approx P(w_2)$ nos equivocamos en promedio uno de cada dos.

$$\boxed{P(error) = min\ [P(w_1), P(w_2)]}$$

DENSIDAD DE PROBABILIDAD CONDICIONADA A LA CLASE

- En general disponemos de más información para tomar decisiones.
- Ejemplo: a cada pixel le asociamos un vector $\mathbf{x}=(x_1,x_2,x_3,x_4)$ donde x_i : reflectancia en la banda espectral i-ésima.
- Modelo: $\mathbf{x}$ vector aleatorio

  $p(\mathbf{x}/w_i)$ densidad de probabilidad

$$\forall R \subset R^d, \quad \int_R p(\mathbf{x}/w_i)\,d\mathbf{x} = P(\mathbf{x} \in R / w_i)$$

BAYES

Supuesto conocidas las prioris y las densidades condicionales

Para inferir la naturaleza del pixel de vector de características $\mathbf{x}$, usamos Bayes : $p(\mathbf{x}, w_i) = P(w_i / \mathbf{x})\, p(\mathbf{x}) = p(\mathbf{x} / w_i) P(w_i)$

$$\Rightarrow \quad \boxed{P(w_i / \mathbf{x}) = \frac{p(\mathbf{x} / w_i) P(w_i)}{p(\mathbf{x})}}$$

donde $$p(\mathbf{x}) = \sum_{i=1}^{2} P(\mathbf{x} / w_i) P(w_i)$$

BAYES

$$\text{Bayes}: \quad posterior = \frac{verosimilitud \times prior}{evidencia}$$

- *$P(w_i/\mathbf{x})$- posterior:* probabilidad de que la clase sea w_i dado que se midió $\mathbf{x}$.
- *$P(w_i)$ – prior:* conocimiento previo del problema
- *$p(\mathbf{x}/w_i)$- verosimilitud* : de la clase w_i respecto a $\mathbf{x}$, cuanto mayor más probable que la verdadera clase sea w_i.
- *$p(\mathbf{x})$- evidencia:* factor de escala, normaliza a 1.

REGLA DE DECISIÓN DE BAYES

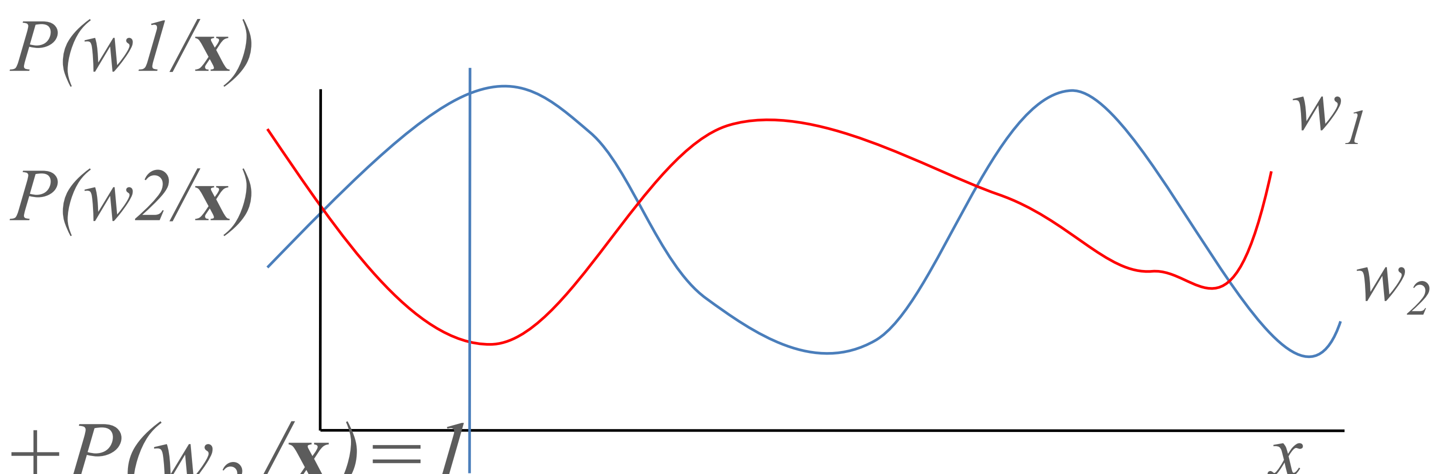
$$\text{Decido } w_1 \text{ si } P(w_1/\mathbf{x}) > P(w_2/\mathbf{x}),\ w_2 \text{ en otro caso}$$

$$P(error) = \int_{R^4} P(error,\mathbf{x})d\mathbf{x} = \int_{R^4} P(error/\mathbf{x})\,p(\mathbf{x})d\mathbf{x}$$

$$P(error/\mathbf{x}) = \begin{cases} P(w_1/\mathbf{x}) & \text{si decidimos } w_2 \\ P(w_2/\mathbf{x}) & \text{si decidimos } w_1 \end{cases}$$

$$P(error)\ \text{mínima} \Leftrightarrow P(error/\mathbf{x})\ \text{mínima}\ \forall\mathbf{x}$$

$$\text{Bajo esta regla}: P(error/\mathbf{x}) = \min\left[P(w_1/\mathbf{x}), P(w_2/\mathbf{x})\right]$$



$$P(w_1/\mathbf{x})+P(w_2/\mathbf{x})=1$$

En término de probabilidad a priori y verosimilitud la regla es:

$$\boxed{\textit{Decido } w_1 \textit{ si } p(\mathbf{x}/w_1)\,P(w_1) > p(\mathbf{x}/w_2)\,P(w_2)\textit{ , } w_2 \textit{ en otro caso}}$$

Eliminando el factor de escala se obtiene una regla equivalente, el factor de normalización cambia la apariencia de las funciones discriminantes.

- Si $p(\mathbf{x}/ w_1)=p(\mathbf{x}/ w_2)$ entonces el medir las características **x,** no nos aporta información sobre la clase; la decisión se basa puramente en las priors.
- Si $P(w_1) =P(w_2)$ la decisión se basa en las verosimilitudes
- La regla de decisión bayesiana combina ambos factores y toma la decisión que minimiza la probabilidad de error.

FORMALIZACIÓN Y GENERALIZACIÓN

- c clases $\{w_1, w_2 \dots . w_c\}$
- Espacios de características R^d :

  $\mathbf{x} = (x_1, x_2, \dots x_d)^T$: vector de características

  $\mathbf{x} \in R^d$: espacio Euclideano dimensión d
- Funciones de costos más generales que la probabilidad de error. Función de costo ó de pérdida: cuanto me cuesta cometer distinto tipos de errores o no decidir.
- Ej: costo de extraer tejido si es benigno no es igual costo de no extraer tumor maligno.

FUNCIONES DE COSTO

$C = \{w_1, w_2 \ldots. w_c\}$ conjunto finito de c categorías

$A = \{\alpha_1, \alpha_2 \ldots. \alpha_a\}$ conjunto de a posibles acciones

$\lambda(\alpha_i / w_j)$: costo asociado a tomar acción α_i, cuando la muestra es en realidad de la clase wj.

RIESGO CONDICIONAL

$$\text{Modelo}: \quad \mathbf{x} \in R^d$$

$$\Omega \in C, \text{ variable aleatoria}$$

$$P(w_j / \mathbf{x}) = \frac{p(\mathbf{x}/w_j)P(w_j)}{p(\mathbf{x})} \qquad p(\mathbf{x}) = \sum_{i=1}^{c} p(\mathbf{x}/w_j)P(w_j)$$

Observo $\mathbf{x}$, contemplo tomar acción α_i, si clase verdadera w_j, voy a incurrir en un costo $\lambda(\alpha_i / w_j)$

El costo medio de tomar la acción α_i es el riesgo condicional:

$$\boxed{R(\alpha_i / \mathbf{x}) = E_{\Omega/\mathbf{x}}\left[\lambda(\alpha_i / \Omega)\right] = \sum_{j=1}^{c} \lambda(\alpha_i / w_j) P(w_j / \mathbf{x})}$$

RIESGO TOTAL

- Una regla de decisión es una función

  $\alpha(\mathbf{x}) \quad \alpha: R^d \rightarrow A$

  que nos dice que acción tomar para cada **x**
- Ej: asigno a una de las clases (1...c) o a la clase de rechazo.
- Riesgo total R: esperanza del riesgo condicional asociado a una regla de decisión

$$\boxed{R = E_{\mathbf{x}}(R(\alpha(\mathbf{x})/\mathbf{x})) = \int_{R^d} R(\alpha(\mathbf{x})/\mathbf{x})\, p(\mathbf{x})\, d\mathbf{x}}$$

RIESGO DE BAYES

- Elegir Regla de Decisión que minimice Riesgo Total

R es mínimo $\leftrightarrow$ para cada $\mathbf{x}$ $R(\alpha(\mathbf{x})/\mathbf{x})$ es mínimo

$\leftrightarrow$ para cada $\mathbf{x}$: $\alpha^{*}(\mathbf{x}) = \arg\min_{\alpha_i \in A} R(\alpha_i/\mathbf{x})$

El riesgo asociado a la Decisión Bayesiana :

$$R^{*} = \int_{R^d} R(\alpha^{*}/\mathbf{x})\, p(\mathbf{x})\, d\mathbf{x}$$

Es ópitmo. En caso de empate entre reglas se puede tomar cualquiera

CLASIFICACIÓN CON 2 CLASES (SIN RECHAZO)

$$\lambda_{ij} = \lambda(\alpha_i / w_j) \qquad C = \{w_1, w_2\} \qquad A = \{\alpha_1, \alpha_2\}$$

$$R(\alpha_1 / \mathbf{x}) = \lambda_{11} P(w_1 / \mathbf{x}) + \lambda_{12} P(w_2 / \mathbf{x})$$

$$R(\alpha_2 / \mathbf{x}) = \lambda_{21} P(w_1 / \mathbf{x}) + \lambda_{22} P(w_2 / \mathbf{x})$$

Regla Bayesiana:

$$\boxed{\text{Decido}: \ w1 \ \text{si} \ \ R(\alpha_1 / \mathbf{x}) < R(\alpha_2 / \mathbf{x}), \ w2 \ \text{en otro caso}}$$

RAZÓN DE VEROSIMILITUD

$$\Leftrightarrow \quad (\lambda_{12} - \lambda_{22}) P(w_2 / \mathbf{x}) < (\lambda_{21} - \lambda_{11}) P(w_1 / \mathbf{x})$$

$$\Leftrightarrow \quad \boxed{\frac{p(\mathbf{x}/w_1)}{p(\mathbf{x}/w_2)} \underset{w_2}{\overset{w_1}{\gtrless}} \left[\frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}}\right] \frac{P(w_2)}{P(w_1)}}$$

En general $\lambda_{21} > \lambda_{11}$ y $\lambda_{12} > \lambda_{22}$ costo de errar mayor al de acertar.

Interpretación : Decido w_1 cuando la razón de verosimilitud supera un umbral que es independiente de $\mathbf{x}$.

CLASIFICACIÓN DE MENOR TASA DE ERROR:

Función de costo simétrica : Costo cero/uno

$$\forall\ (w_1, w_2) \in C^2$$

$$\lambda(\alpha_i / w_j)\begin{cases} 0 & i = j \ \text{ aciertos} \quad \text{i, j} = 1 \ldots \text{c} \\ 1 & i \neq j \ \text{ errores} \end{cases}$$

Asumo que todos los errores cuestan lo mismo :

$$\boxed{R(\alpha_i / \mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i / w_j) P(w_j / \mathbf{x}) = \sum_{j \neq i} P(w_j / \mathbf{x}) = 1 - P(w_i / \mathbf{x})}$$

CLASIFICACIÓN CON TASA DE ERROR MÍNIMA

- La decisión bayesiana es aquella que minimiza el riesgo total y por ende el riesgo condicional para

  $R(\alpha_i / \mathbf{x})$ para todo $\mathbf{x}$ $\leftrightarrow$

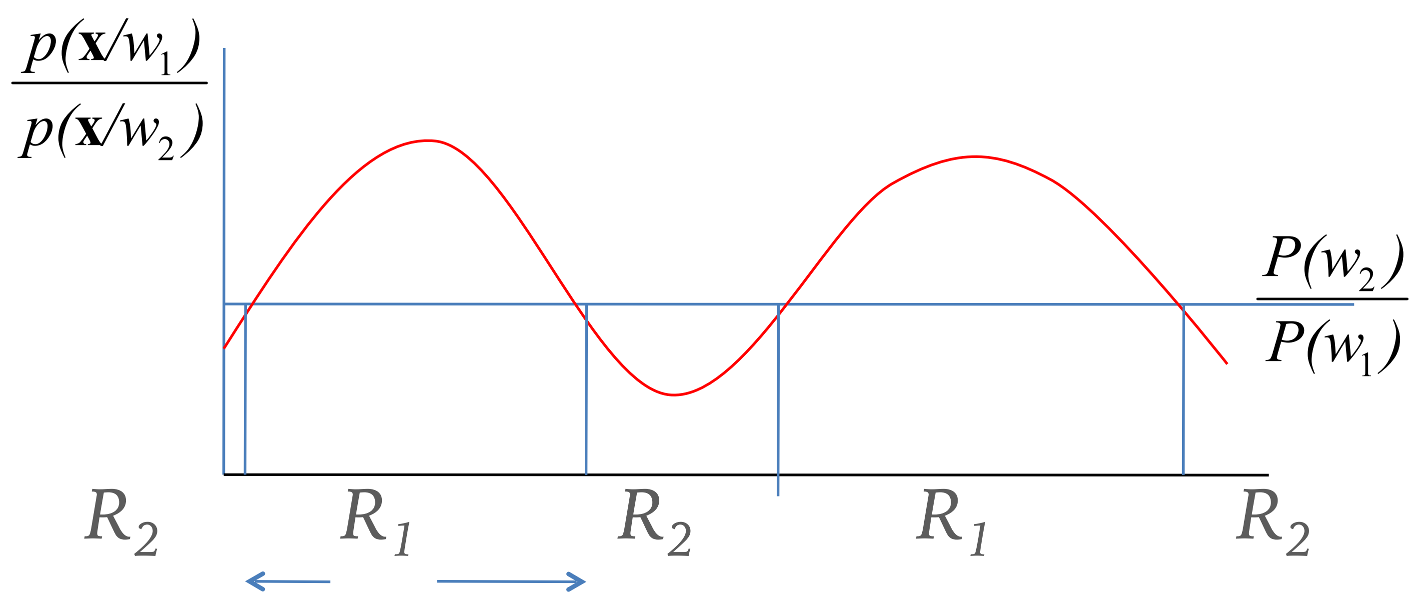
$$\boxed{\text{Decido } w_i \text{ si } P(w_i / x) > P(w_j / x) \text{ para todo } j \neq i}$$

- $P(w_i / \mathbf{x})$ es la probabilidad condicional de que la acción α_i es correcta. Para minimizar el riesgo tengo que elegir i que maximiza $P(w_i / \mathbf{x})$.

REGLA DE DECISIÓN BAYESIANA

$$\lambda_{12} = \lambda_{21} = 1$$

$$\lambda_{11} = \lambda_{22} = 0$$



$$\frac{p(\mathbf{x}/w_1)}{p(\mathbf{x}/w_2)} \underset{w_2}{\overset{w_1}{\gtrless}} \frac{P(w_2)}{P(w_1)}$$

R_i: región donde se decide w_i

No tiene porque ser conexa.

EJERCICIO

- La incidencia por cancer de próstata en el uruguay es de aproximadamente 50 nuevos casos por año cada 100.000 pobladores. Asumimos que, para una institución de salud, la relación entre el costo de no tratar al paciente enfermo y de tratar a una persona sana es de 10.000 a uno. En una campaña de diagnóstico masivo, se somete a un individuo a un test que indica, con incertidumbres gaussianas de varianza 1: cero si el paciente es sano, 8 si no lo es. Pepito se somete al examen y el resultado es 3.9.
- Qué haría la institución?

REPRESENTACIÓN CLASIFICADOR BAYESIANO

Caso genérico : $g_i(\mathbf{x}) = -R(\alpha_i / \mathbf{x})$

$\forall f : R \to R$ monótona creciente $\{g_i\}_{i=1}^{c}$ y $\{f(g_i)\}$ conducen a la misma clasificación. La idea es elegir f para simplificar analítica o computacionalmente.

Ej : $g_i(\mathbf{x}) = P(w_i / \mathbf{x}) = \dfrac{p(\mathbf{x} / w_i) P(w_i)}{p(\mathbf{x})}$ tasa mínima de error

$$g_i(\mathbf{x}) = p(\mathbf{x} / w_i) P(w_i)$$

$$\boxed{g_i(\mathbf{x}) = \ln p(\mathbf{x} / w_i) + \ln P(w_i)}$$

DISTRIBUCIÓN NORMAL

- Clasificación basada en modelos estadísticos determinados por momentos de primer y segundo orden.
- Problema práctico descripto por conjunto de entrenamiento $\{\mathbf{x}, w\}$, no tenemos conocimiento de las propiedades estocásticas de la fuente de patrones.
- Enfoque pragmático: Modelar $p(\mathbf{x}/w)$ usando distribución normal y evaluamos si hipótesis es sensible.
- Encontrar clasificador óptimo para $p(\mathbf{x}/w)$ normal.

DENSIDAD NORMAL MULTIVARIADA

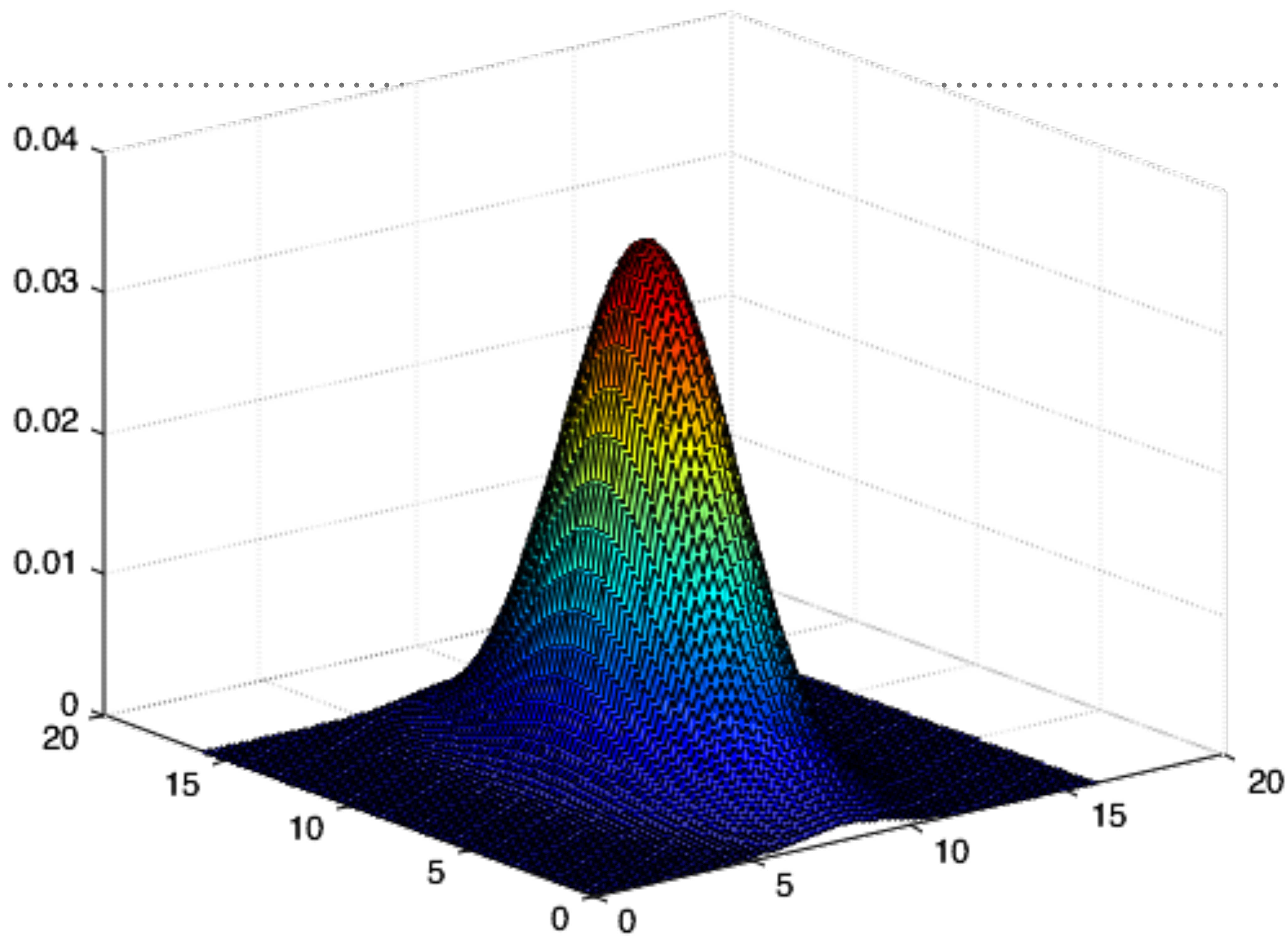
$$\mathbf{x} = \begin{bmatrix} x_1 \\ \\ x_d \end{bmatrix} \quad p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}(\det \boldsymbol{\Sigma})^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu}) \right]$$

$$\boldsymbol{\mu} = E(\mathbf{x}) = \int_{R^d} \mathbf{x} p(\mathbf{x}) d\mathbf{x} \quad \text{Valor medio de } \mathbf{x}$$

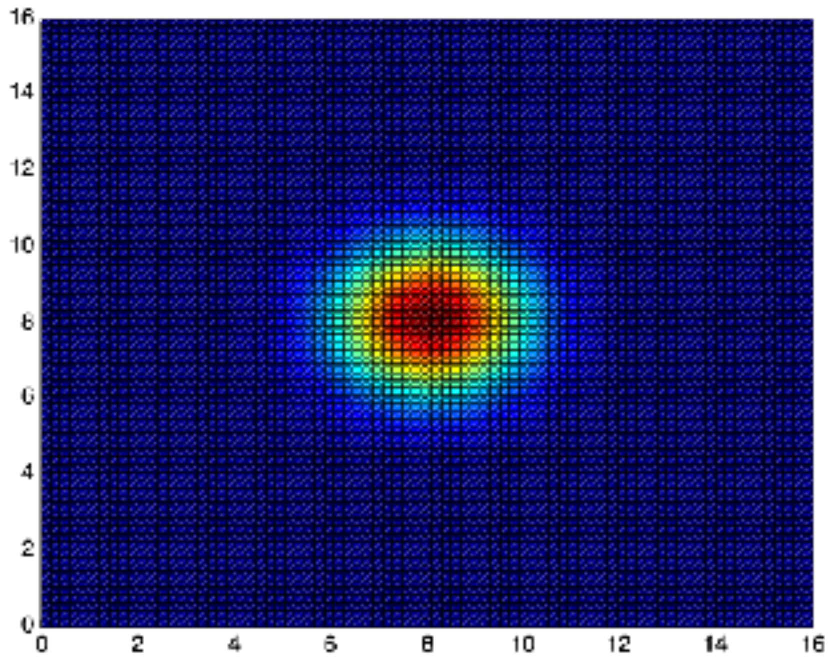
$$\boldsymbol{\Sigma} = E\left[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T \right] = \int_{R^d} (\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T p(\mathbf{x}) d\mathbf{x}$$

$\boldsymbol{\Sigma}$: Matriz de covarianza de $\mathbf{x}$

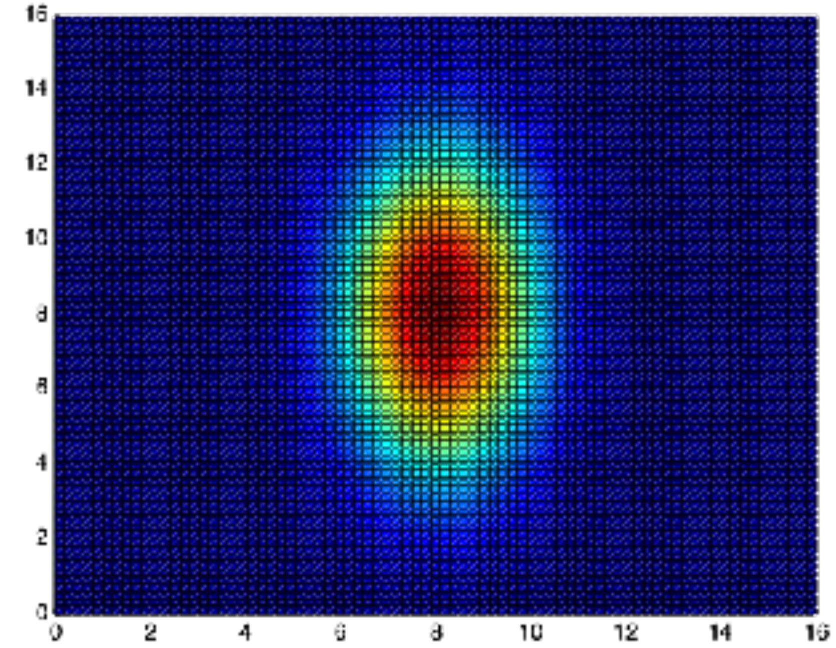
$$r^2 = (\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu}) \quad \text{Distancia cuadrática de Mahalanobis}$$



0.04
0.03
0.02
0.01
0
20
15
10
5
0
0
5
10
15
20

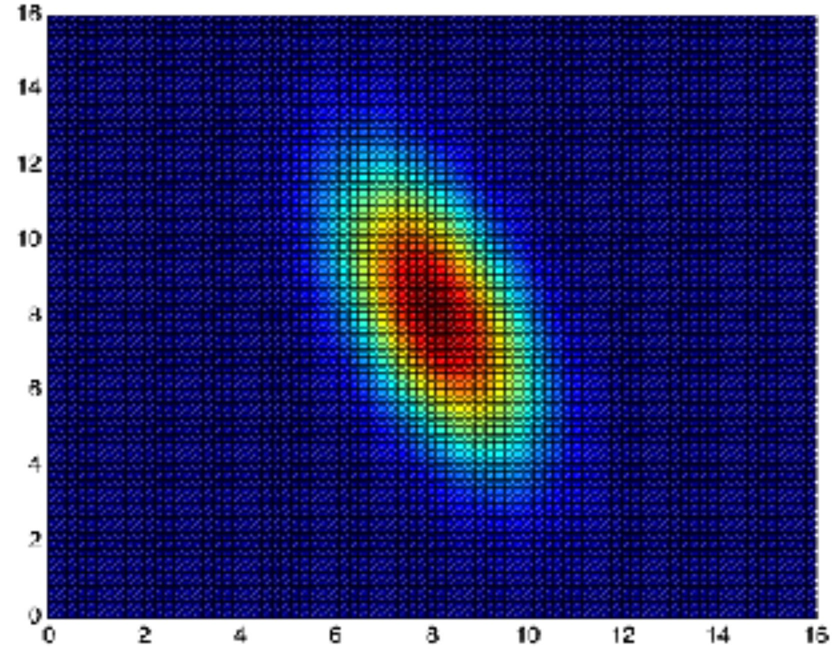


$$\boldsymbol{\Sigma} = \sigma^2 I = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$$



$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11}{}^2 & 0 \\ 0 & \sigma_{22}{}^2 \end{bmatrix}$$

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



ORIENTACIÓN Y TAMAÑO DEL ELIPSOIDE

$\mathbf{B}$: matriz de vectores propios de $\mathbf{\Sigma}$

$\mathbf{D}$: matriz diagonal de valores propios de $\mathbf{\Sigma}$

$\mathbf{\Sigma} = \mathbf{B}\mathbf{D}\mathbf{B}^T$

Cambio de coordenadas $\mathbf{z} = \mathbf{B}^T(\mathbf{x} - \boldsymbol{\mu})$, transformación compuesta traslación (origen de coordenadas a $\boldsymbol{\mu}$) y rotación por la matríz $\mathbf{B}$.

$\rightarrow r^2 = \mathbf{z}^T\mathbf{D}^{-1}\mathbf{z}$ $\qquad r^2 = \sum_1^d \frac{\mathbf{z}_i^{\,2}}{\lambda_i}$ $\qquad \lambda_i$: valores propios

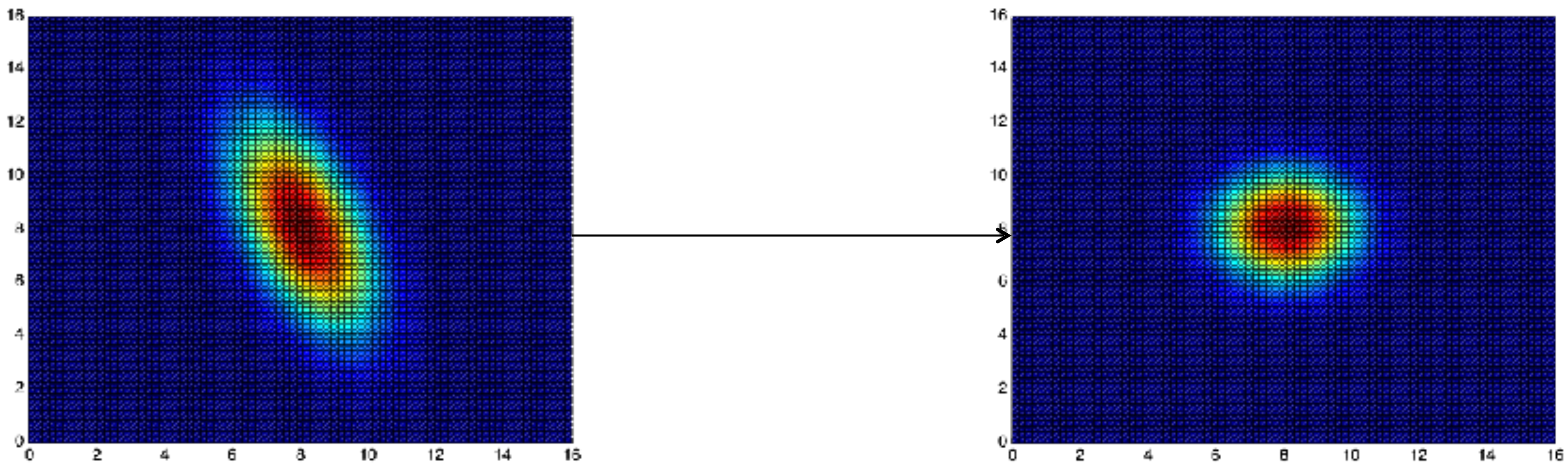
Los ejes del elipsoide estan definidos por los $\mathbf{b}_{\mathrm{i}}$ y la longitud de los semiejes por $\sqrt{\lambda_{\mathrm{i}}}$.

TRANSFORMACIONES DE BLANQUEADO

Transformación de coordenadas que lleva una distribución normal arbitraria en una distribución esférica (blanca $\mathbf{\Sigma} = \sigma\mathbf{I}$)

$$A_w = \mathbf{B}\mathbf{D}^{-1/2}$$

Ej : Blanqueado usando $\mathbf{\Sigma} = \mathbf{B}\mathbf{D}\mathbf{B}^T$ y Cholesky : $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T$



FUNCIONES DISCRIMINANTES CUADRÁTICAS

Funciones discriminantes para clasificar con menor tasa de error

$$g_i(\mathbf{x}) = \ln P(\mathbf{x} / w_i) + \ln P(w_i)$$

Si todas las densidades son normales :

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) - \cancel{\frac{d}{2}\ln 2\pi} - \frac{1}{2}\ln(\det \Sigma_i) + \ln P(w_i)$$

MATRICES DE COVARIANZA BLANCAS

Caso I) $\boxed{\boldsymbol{\Sigma}_{\mathrm{i}} = \sigma^2 \mathbf{I}}$

$$g_i(\mathbf{x}) = -\frac{\left\|\mathbf{x}-\boldsymbol{\mu}_i\right\|^2}{2\sigma^2} + \ln P(w_i) - \frac{1}{2}\cancel{\ln(\sigma^{2d})}$$

Estas funciones no son lineales pero se pueden hacer :

$$(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i) = \mathbf{x}^T\mathbf{x} - 2\boldsymbol{\mu}_i^T\mathbf{x} + \boldsymbol{\mu}_i^T\boldsymbol{\mu}_i$$

$\mathbf{x}^T\mathbf{x}$: podemos ignorarlo (igual $\forall\, g_i$)

$$g_i(\mathbf{x}) = \frac{1}{\sigma^2}\boldsymbol{\mu}_i^T\mathbf{x} + \ln P(w_i) - \frac{1}{2\sigma^2}\boldsymbol{\mu}_i^T\boldsymbol{\mu}_i$$

$$\leftarrow \mathbf{w}_i^T \rightarrow \qquad \leftarrow \quad \mathbf{w}_{i0} \quad \rightarrow$$

$$\boxed{g_i(\mathbf{x}) = \mathbf{w}_i^T\mathbf{x} + w_{i0}}$$

Modelo adecuado ruido blanco gaussiano no correlacionado independiente de la clase con varianza $\sigma = \mathrm{var}(r) = \mathrm{var}(\mathbf{x})$ superpuesto a los vectores prototipos μ_k

$w_i \propto \mu_i$ (filtro apareado)

SUPERFICIES DE DECISIÓN

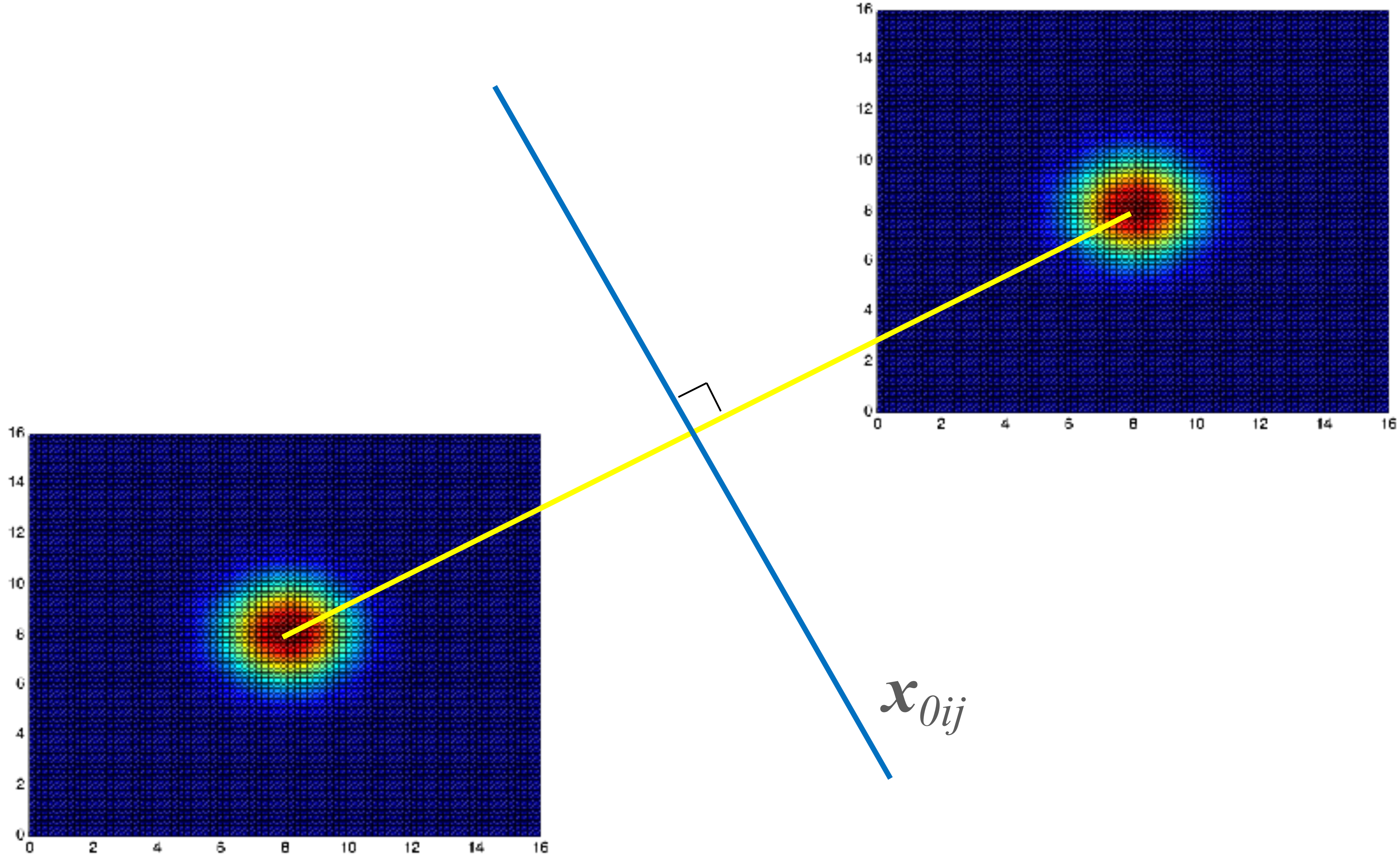
$$s_{ij} = \{\mathbf{x} / g_i(\mathbf{x}) = g_j(\mathbf{x})\}$$

$$s_{ij} = \{\mathbf{x} / (\mathbf{w}_i - \mathbf{w}_j)^T \mathbf{x} + (w_{i0} - w_{j0}) = 0\}$$

Reordenando :

$$0 = \underbrace{\frac{1}{\sigma^2}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T}_{\mathbf{w}_{ij}^T} \left[\mathbf{x} + \underbrace{\sigma^2 \frac{(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)}{\left\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right\|^2} \ln \frac{P(w_i)}{P(w_j)} - \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j)}_{-\mathbf{x}_{0ij}}\right]$$

$$\boxed{s_{ij} = \{\mathbf{x} / \mathbf{w}_{ij}^T (\mathbf{x} - \mathbf{x}_{0ij}) = 0\}}$$



0
2
4
6
8
10
12
14
16
x_{0ij}

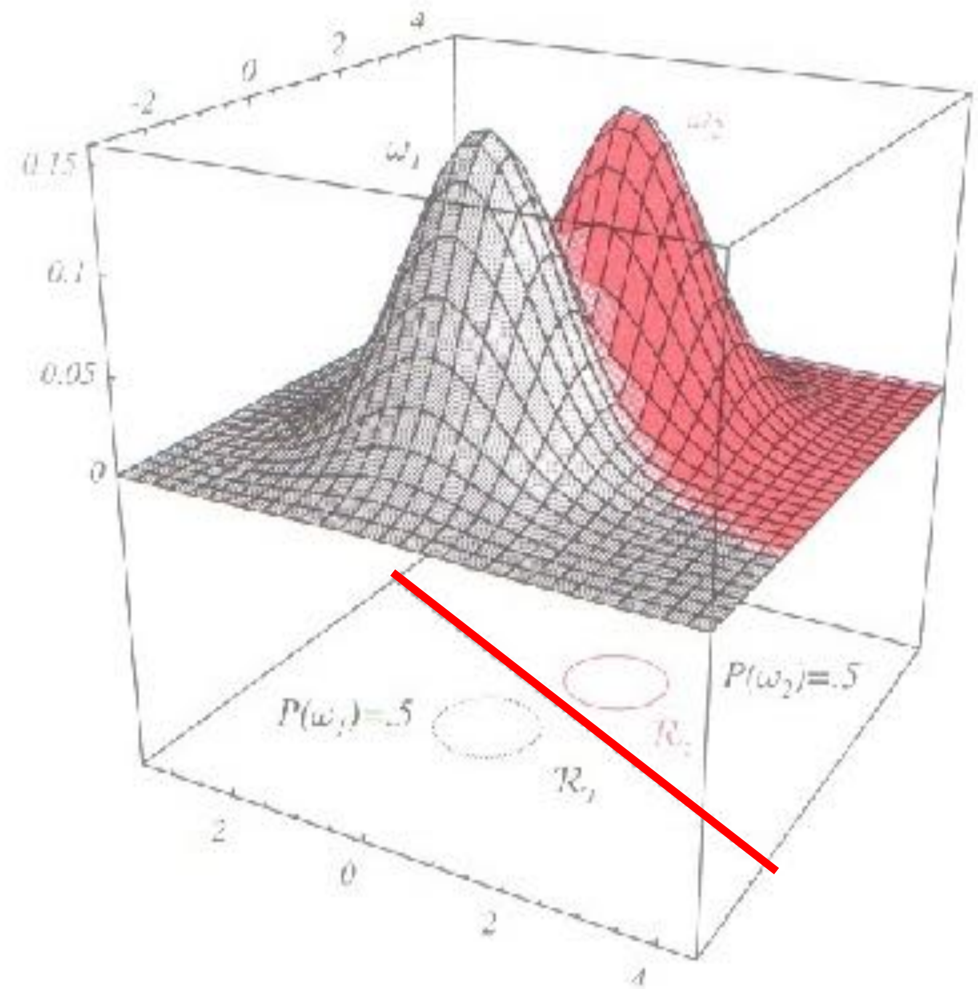
SUPERFICIES DE DECISIÓN

Si $P(w_i) = P(w_j) \rightarrow \mathbf{x}_{0ij} = \frac{1}{2}\left(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j\right)$

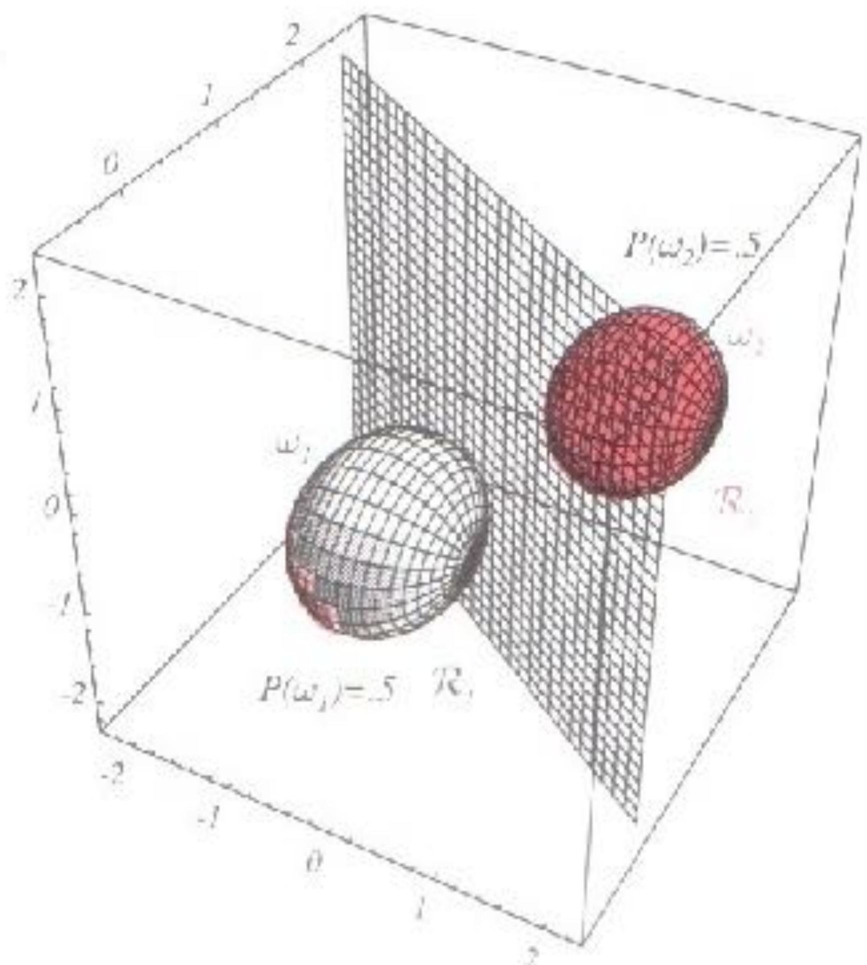
Si $P(w_i) > P(w_j) \rightarrow \mathbf{x}_{0ij}$ se acerca a $\boldsymbol{\mu}_j$

Si $\sigma^2 << \left\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right\| \rightarrow s_{ij}$ menos sensibles a los prioris

(los datos tienen poca incertidumbre $\rightarrow$ creemos más en las observaciones)



0.15
0.1
0.05
0
ω_1
ω_2
$P(\omega_1)$=.5
$P(\omega_2)$=.5
$\mathcal{R}_1$
$\mathcal{R}_2$



$P(\omega_2)$=.5
$P(\omega_1)$=.5
ω_1
ω_2
$\mathcal{R}_1$
$\mathcal{R}_2$

Duda

CLASIFICADOR DE MÍNIMA DISTANCIA (EUCLIDEO)

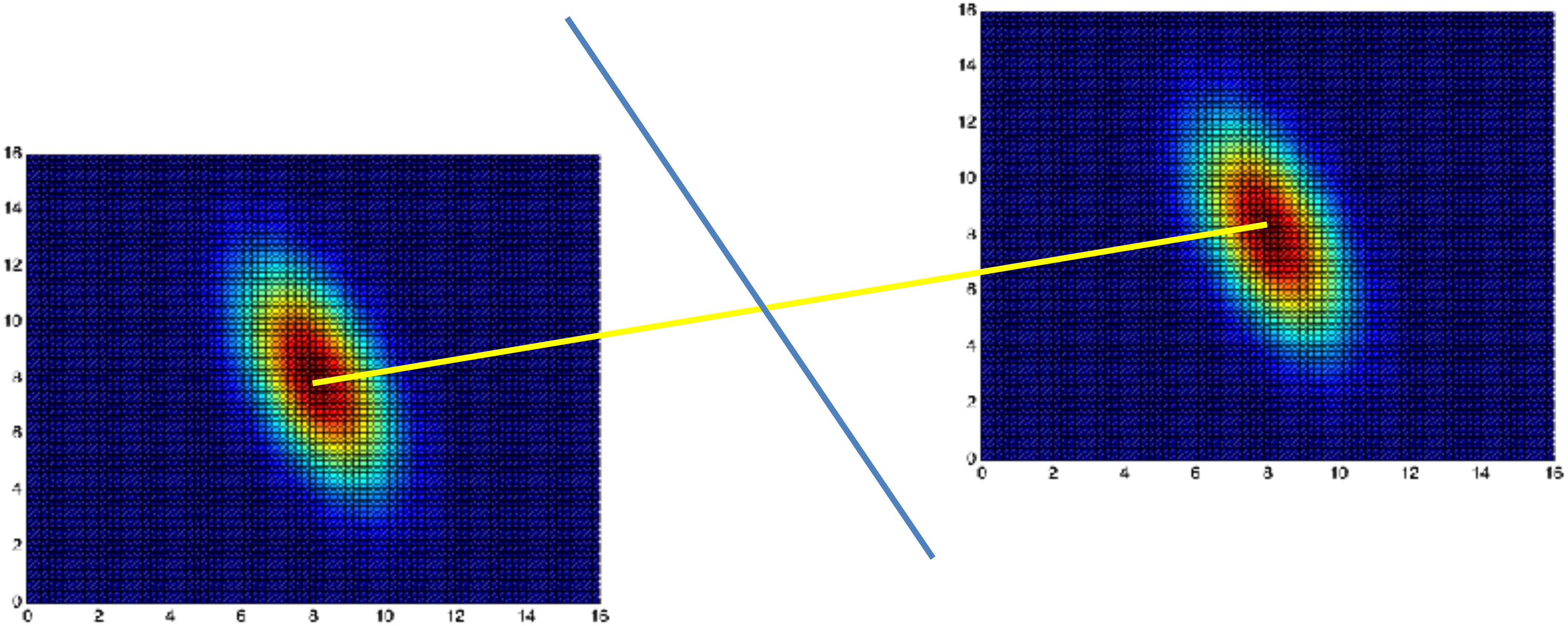
$\Sigma = \sigma \mathrm{I}$ - Si los prioris son equiprobables o casi $\ln \dfrac{P(w_i)}{P(w_j)} \approx 0$

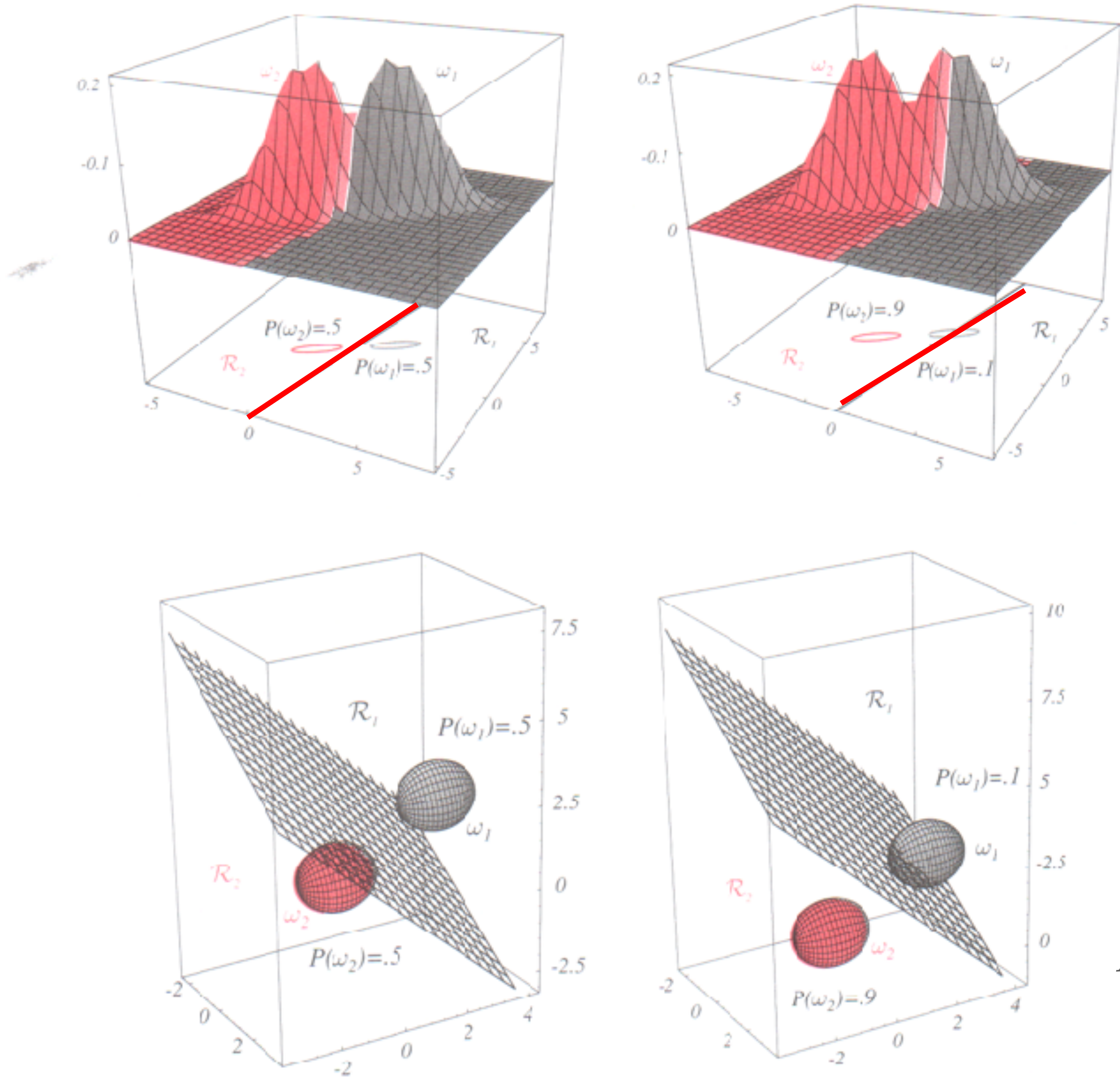
$$\left(P(w_i) \approx \frac{1}{c} \quad \forall i = 1, \ldots c\right)$$

$$\boxed{\text{Decido: Asignar } \mathbf{x} \text{ a la clase } w_i^* \,/\, i^* = \arg\min_{i=1..c} \left\| \mathbf{x} - \boldsymbol{\mu}_i \right\|}$$

El clasificador euclideo es óptimo para clases con distribución normal con matrices de covarianza iguales y proporcionales a la identidad y prioris equiprobables.

A menos que $(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$ sea vector propio de $\boldsymbol{\Sigma}$ el hiperplano de separación no va ser ortogonal a $(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$.





Duda

CLASIFICADOR DE MAHALANOBIS (MINIMA DISTANCIA)

$\Sigma_i = \Sigma$ - Si los prioris son unifor mes o casi $\ln \dfrac{P(w_i)}{P(w_j)} \approx 0$

$$\left(\frac{P(w_i)}{P(w_j)} \approx \frac{1}{c} \quad \forall i = 1,\ldots..c\right)$$

$$\mathrm{g_i}(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma (\mathbf{x} - \boldsymbol{\mu}_i) = r^2$$

$$\boxed{\text{Decido : Asignar } \mathbf{x} \text{ a la clase } w_i^* \,/\, i^* = \underset{i=1..c}{\arg\min}\, g_i(x)}$$

El clasificador de Mahalanobis es óptimo para clases con distribución normal con matrices de covarianza iguales y prioris equiprobables.

NAIVE BAYES

- Cuando no se conocen las relaciones de dependencia entre las características, se asume lo más simple, independencia. En la práctica puede funcionar bien a pesar de su simplicidad.

$$p(\omega_k|\mathbf{x}) \propto \prod_{i=1}^{d} p(x_i|\omega_k)$$

```
>>> from sklearn.naive_bayes import GaussianNB
>>> gnb = GaussianNB()
>>> y_pred = gnb.fit(iris.data, iris.target).predict(iris.data)
```

- http://scikit-learn.org/stable/modules/naive_bayes.html