mit:

$$g_0 = \ln\frac{\det \mathbf{K}_i}{\det \mathbf{K}_j} - 2\ln\frac{P(\omega_i)}{P(\omega_j)} + \boldsymbol{\mu}_{x_i}^T \mathbf{K}_i^{-1} \boldsymbol{\mu}_{x_i}$$

$$- \boldsymbol{\mu}_{x_j}^T \mathbf{K}_j^{-1} \boldsymbol{\mu}_{x_j} + \mathbf{x}_0^T \mathbf{M}^{-1} \mathbf{x}_0$$

$$\mathbf{x}_0 = \mathbf{M}[\mathbf{K}_i^{-1}\boldsymbol{\mu}_{x_i} - \mathbf{K}_j^{-1}\boldsymbol{\mu}_{x_j}]$$

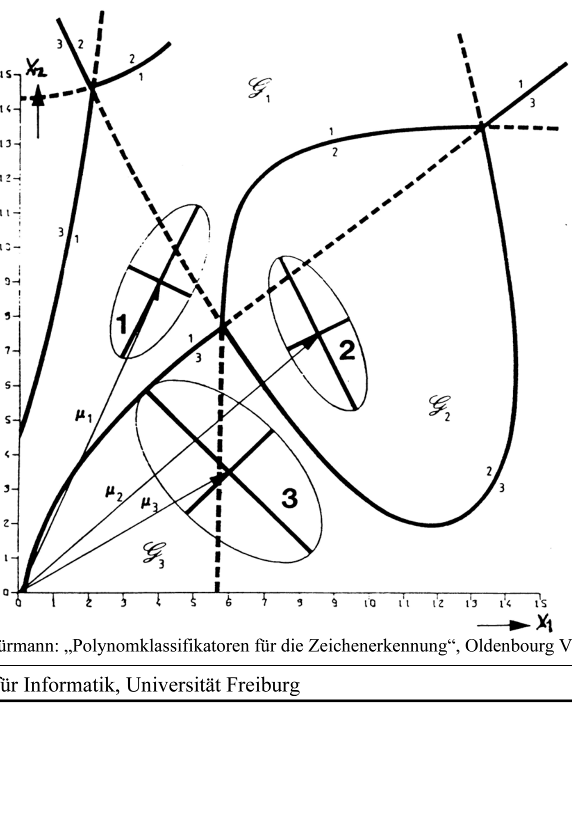
$$\mathbf{M} = [\mathbf{K}_i^{-1} - \mathbf{K}_j^{-1}]^{-1} = \mathbf{K}_i[\mathbf{K}_j - \mathbf{K}_i]^{-1}\mathbf{K}_j$$

$$= \mathbf{K}_j[\mathbf{K}_j - \mathbf{K}_i]^{-1}\mathbf{K}_i$$

Die die quadratische Form charakterisierende Matrix $\mathbf{M}^{-1}$ ist nun nicht mehr zwingend pos. Definit => die Grenzflächen zwischen den Gebieten sind allgemeine Kegelschnitte (bei N=2: Ellipsen, Parabeln, Hyperbeln, Linien)

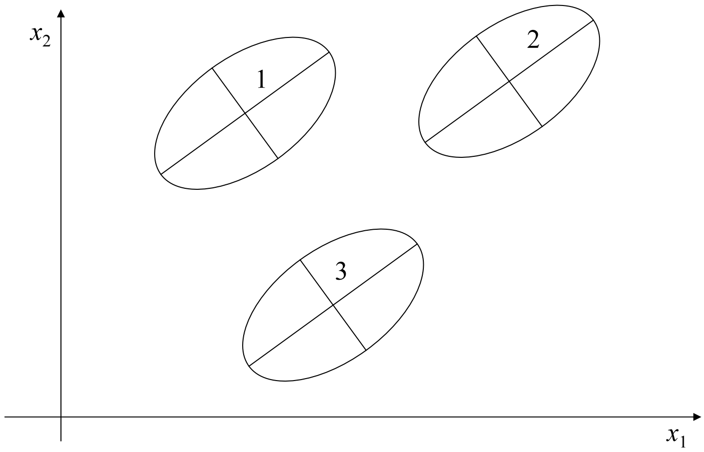
Die Unterscheidungsfunktionen $D'_k(\mathbf{x})$ sind in Bezug auf den Merkmalsvektor quadratische Funktionen oder Polynome zweiten Grades (***quadratischer oder Polynomklassifikator***)

Klassenweise normalverteilte Merkmale



(aus J. Schürmann: „Polynomklassifikatoren für die Zeichenerkennung“, Oldenbourg Verlag)

Klassenweise normalverteilte Merkmale mit identischen Kovarianzmatrizen **K**



Klassenweise normalverteilte Merkmale mit identischen Kovarianzmatrizen **K**

Entscheidungsfunktion:

$$D'_k(\mathbf{x}) = \ln P(\omega_k) - \frac{1}{2}\ln(\det \mathbf{K}) - \frac{1}{2}[(\mathbf{x} - \boldsymbol{\mu}_{x_k})^T \mathbf{K}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{x_k})]$$

$$D''_k = -2D'_k - \ln(\det \mathbf{K})$$

$$\Rightarrow D''_k = -2\ln P(\omega_k) + (\mathbf{x} - \boldsymbol{\mu}_{x_k})^T \mathbf{K}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{x_k})$$

Für gleiche Auftretenswahrscheinlichkeiten $P(\omega_k) = 1/K$ folgt:

$$D'''_k = D''_k - 2\ln K$$

$$\Rightarrow \boxed{D'''_k = d_M^2 = (\mathbf{x} - \boldsymbol{\mu}_{x_k})^T \mathbf{K}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{x_k})}$$ ***Mahalanobis-Abstands-Klassifiaktor***

Dies ist eine allgemeine gewichtete quadratische Metrik.

Formulierung als linearer Klassifikator

Die Entscheidungsfunktion enthält noch einen quadratischen Term. Dieser ist jedoch für jede Klasse identisch und kann somit eliminiert werden. Damit kommt man zu einer linearen Formulierung des Klassifikators.

Alternativ ergibt sich:

$$D_k'' = D_k' + \frac{1}{2}[\ln(\det \mathbf{K}) + \mathbf{x}^T \mathbf{K}^{-1}\mathbf{x}]$$

$$\Rightarrow D_k'' = \ln P(\omega_k) - \frac{1}{2}\boldsymbol{\mu}_{x_k}^T \mathbf{K}^{-1}\boldsymbol{\mu}_{x_k} + \boldsymbol{\mu}_{x_k}^T \mathbf{K}^{-1}\mathbf{x}$$

Dieser Ausdruck ist linear in $\mathbf{x}$!

$$\Rightarrow \boxed{D_k''(\mathbf{x}) = a_{0k} + \mathbf{a}_k^T \mathbf{x}}$$

Hyperebene als Trennfläche !

mit:

$$a_{0k} = \ln P(\omega_k) - \frac{1}{2}\boldsymbol{\mu}_{x_k}^T \mathbf{K}^{-1}\boldsymbol{\mu}_{x_k}$$

$$\mathbf{a}_k = \mathbf{K}^{-1}\boldsymbol{\mu}_{x_k}$$

Klassenweise normalverteilte Merkmale mit der Einheitsmatrix als Kovarianzmatrix $\mathbf{K}=\sigma^2\mathbf{I}$

(sphärisch invariante Verhältnisse, Hyperkugeln)

Entscheidungsfunktion:

$$D_k'(\mathbf{x}) = \ln P(\omega_k) - N \ln\sigma - \frac{1}{2\sigma^2}(\mathbf{x} - \boldsymbol{\mu}_{x_k})^T(\mathbf{x} - \boldsymbol{\mu}_{x_k})$$

Für konstante A-priori-W. folgt:

$$\Rightarrow \boxed{D_k'' = \left\|\mathbf{x} - \boldsymbol{\mu}_{x_k}\right\|^2}$$

Euklidsche Metrik
Minimum-Abstands-Klassifikator

Auch hier formuliert als linearer Klassifikator:

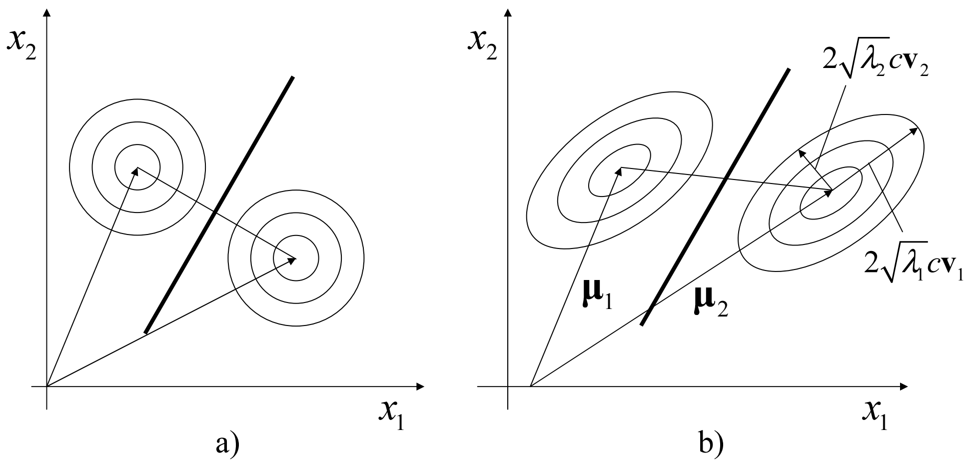
$$D_k''' = \frac{1}{2}(\|\mathbf{x}\|^2 - D_k'')$$

$$\Rightarrow \boxed{D_k'''(\mathbf{x}) = a_{0k} + \mathbf{a}_k^T \mathbf{x}}$$

mit:

$$D_k''' = \frac{1}{2}(\|\mathbf{x}\|^2 - (\|\mathbf{x}\|^2 - \|\boldsymbol{\mu}_{x_k}\|^2 - 2 < \mathbf{x}, \boldsymbol{\mu}_{x_k} >= \underbrace{-\frac{1}{2}\|\boldsymbol{\mu}_{x_k}\|^2}_{a_{0k}} + < \underbrace{\boldsymbol{\mu}_{x_k}}_{\mathbf{a}_k}, \mathbf{x} >$$

Kurven konstanter a) Euklidscher- und b) Mahalanobis-Distanz d_M zum Erwartungswert der jeweiligen Klasse



Transformation der Mahalanobis-Metrik auf sphärisch invariante Verhältnisse

Die Kovarianzmatrix kann mit einer KLT diagonalisiert werden:

$$\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N) = \mathbf{A}^T\mathbf{K}\mathbf{A}$$

bzw:

$$\mathbf{K} = \mathbf{A}\boldsymbol{\Lambda}\mathbf{A}^T$$

Die unitäre Matrix erfüllt: $\mathbf{A}^T = \mathbf{A}^{-1}$

Die Eigenwerte sowie die Eigenvektoren von K definieren die Diagonalmatrix und die Eigenvektoren die Tansformationsmatrix:

$$\mathbf{A} = [\mathbf{e}'_1, \mathbf{e}'_2, \ldots, \mathbf{e}'_N]$$

Kurven konstanter Mahalanobisdistanz ergeben sich zu:

$$(\mathbf{x} - \boldsymbol{\mu}_i)^T \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^T(\mathbf{x} - \boldsymbol{\mu}_i) = c^2$$

Transformation auf sphärisch invariante Verhältnisse

Einführung einer Koordinatentransformation: $\mathbf{x}' = \mathbf{A}^T\mathbf{x}$

Damit werden die ursprünglichen Koordinaten auf die Eigenvektoren projeziert und man kommt man zu den folgenden Kurven konstanter Mahalanobis-Distanz:

$$\frac{(x_1' - \mu_{i1}')^2}{\lambda_1} + \cdots + \frac{(x_N' - \mu_{iN}')^2}{\lambda_N} = c^2$$

Dies ist ein Hyperellipsoid in dem neuen Koordinatensystem.

Mit $x_k'' = x_k' / \lambda_k$ und $\mu_{ik}'' = \mu_{ik}' / \lambda_k$ erhält man sphärisch invariante Verhältnisse:

$$(x_1'' - \mu_{i1}'')^2 + \cdots + (x_N'' - \mu_{iN}'')^2 = c^2 \qquad \text{(Kugeln)}$$

Beispiel: Zweiklassenproblem der Dimension 2

Die Kovarianzmatrix und die Erwartungswerte seien gegeben mit:

$$\mathbf{K} = \begin{bmatrix} 1.1 & 0.3 \\ 0.3 & 1.9 \end{bmatrix}$$

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 0 & 0 \end{bmatrix}^T \quad \text{und} \quad \boldsymbol{\mu}_2 = \begin{bmatrix} 3 & 3 \end{bmatrix}^T$$

Klassifiziere die Beobachtung $\mathbf{x} = \begin{bmatrix} 1.0 & 2.2 \end{bmatrix}^T$ nach Bayes.

Die geschieht durch Berechnung der Mahalanobisdistanz zu den beiden Erwartungswerten:

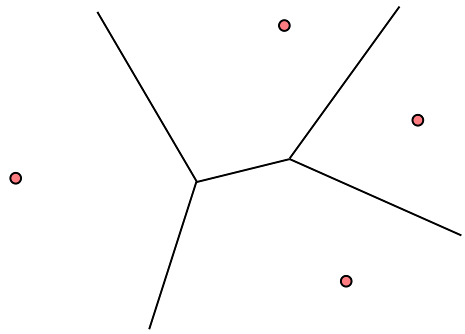
$$d_M^2(\boldsymbol{\mu}_1, \mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_1)^T \mathbf{K}^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) = \begin{bmatrix} 1.0 & 2.2 \end{bmatrix} \begin{bmatrix} 0.95 & -0.15 \\ -0.15 & 0.55 \end{bmatrix} \begin{bmatrix} 1.0 \\ 2.2 \end{bmatrix} = 2.952$$

enstprechend:

$$d_M^2(\boldsymbol{\mu}_2, \mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_2)^T \mathbf{K}^{-1} (\mathbf{x} - \boldsymbol{\mu}_2) = \begin{bmatrix} -2.0 & -0.8 \end{bmatrix} \begin{bmatrix} 0.95 & -0.15 \\ -0.15 & 0.55 \end{bmatrix} \begin{bmatrix} -2.0 \\ -0.8 \end{bmatrix} = 3.672$$

D.h. die Beobachtung wird der Klasse 1 zugeordnet. Man beachte, dass die Beobachtung bzgl. der Euklidschen Distanz näher an Klasse 2 liegt!!

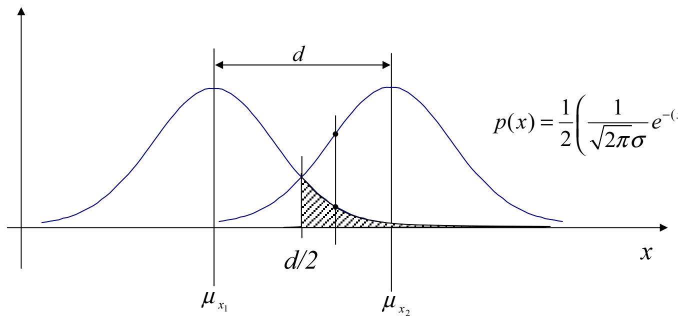
Das Voronoi-Diagramm im zweidimensionalen Raum für die Euklidsche Distanz



Zerlegung der Ebene in Regionen R_i für eine Menge von Punkten. Jede Region enthält genau die Punkte, welche näher sind zu den jeweiligen Punkten als irgend ein anderer Punkt:

$$R_i = \{\mathbf{x} : d(\mathbf{x}, \mathbf{x}_i) < d(\mathbf{x}, \mathbf{x}_j) \quad \text{für} \quad i \neq j\}$$

Wahrscheinlichkeit einer Fehlklassifikation bei einem Zwei-Klassen-Problem



$$p(x) = \frac{1}{2}\left(\frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}\right) = \frac{1}{2} g(x)$$

Die W. dass ω_1 gesendet wurde und bei gemessenen Werten x zugunsten von ω_2 entschieden wird, entspricht der W. dass x oberhalb von von $d/2$ liegt (schraffierte Fläche). Ebenso für $\omega_2 \rightarrow \omega_1$. D.h. die W. einen Fehler bei der Klassifikation zu machen ergibt sich aus der doppelten Fläche:

$$P(E) = 2 \cdot \frac{1}{2} \int_{x=d/2}^{\infty} g(u) du$$

Mit der kumulativen Verteilungsfunktion $F(x_0) = P(x \le x_0)$ gilt:

$$F(x) = \int_{-\infty}^{x} g(u)du = \frac{1}{2}(1 + \mathrm{erf}\left(\frac{x-\mu_x}{\sqrt{2}\sigma}\right))$$

und mit der Gauß'schen Fehlerfunktion

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

bzw. komplementären Fehlerfunktion

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt = 1 - \mathrm{erf}(x)$$

erhält man die Fehler-W. zu:

$$P(E) = \int_{x=d/2}^{\infty} g(u)du = 1 - F(x = d/2)$$

$$= \frac{1}{2}(1 - \mathrm{erf}\left(\frac{x-\mu_x}{\sqrt{2}\sigma}\right))\Bigg|_{\substack{x=d/2\\ \mu_x=0}} = \frac{1}{2}\mathrm{erfc}\left(\frac{x-\mu_x}{\sqrt{2}\sigma}\right)\Bigg|_{\substack{x=d/2\\ \mu_x=0}}$$

$$\boxed{P(E) = \frac{1}{2}\mathrm{erfc}\left(\frac{d/2}{\sqrt{2}\sigma}\right)}$$

Die W. für eine Fehlklassifikation sinkt mit wachsendem Klassenabstand und steigt mit wachsender Streuung der Merkmale

Die Gesamtfehlerwahrscheinlichkeit im N-dimensionalen Merkmalsraum berechnet sich zu (Forney):

$$\boxed{P(E) = const \cdot \frac{1}{2}\mathrm{erfc}\left(\frac{d/2}{\sqrt{2}\sigma}\right)}$$

(ohne Beweis)