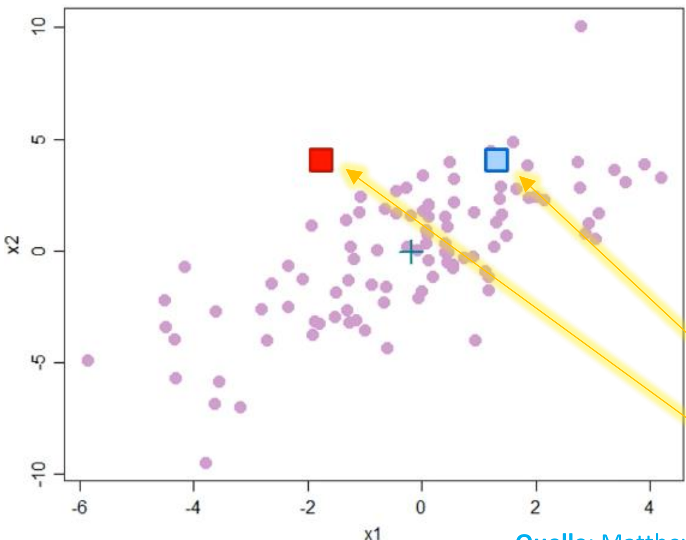
Aufgabenstellungen:
auch unten auf den Folien, vor jeweiligen Lösungen

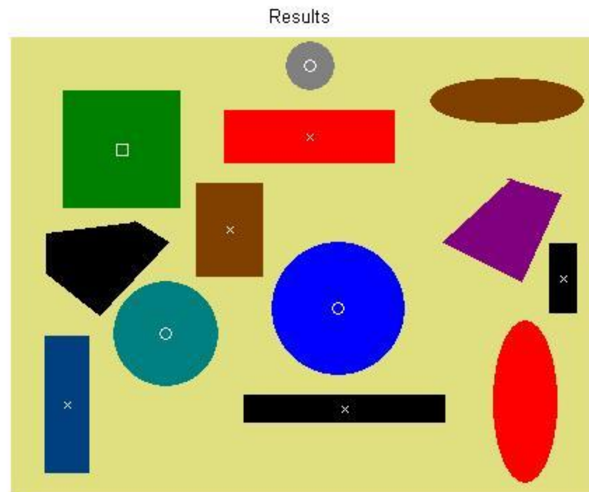
Übungsblatt_06	*http://www.uni-forst.gwdg.de/~wkurth/bia20_ue06.pdf*
Übungsblatt_04	*http://www.uni-forst.gwdg.de/~wkurth/bia20_ue04.pdf*
Übungen_06	*http://www.uni-forst.gwdg.de/~wkurth/bia20_ub08.pdf*

Die Themenpalette in der heutigen Sitzung:

Einteilung von Objekten in Klassen	***Einführung in die Klassifikationsverfahren***
Anhand gebräuchlichsten Abstandsklassifikatoren	***Erkennen von geometrischen Figuren***



Quelle: Matthew Clapham



Quelle: MathWorks/MATLAB

Interessanter Einstieg: die beiden durch ein kleines Quadrat gekennzeichneten Punkte haben den gleichen Abstand zum Datenschwerpunkt (**blaues Kreuzchen**) – „euklidisch“ ☺ gesehen sind diese nicht zu unterscheiden.

Vom statistischen Standpunkt aus betrachtet, sind diese Punkte aber etwas unterschiedlich, da einer (der rote) eher einen Ausreißer darstellt.

Genau um solche Momente in Betracht zu ziehen, hat man **Euklidische Distanz** etwas erweitert - so lässt sich die **statistische Distanz** intuitiverweise verstehen.

Inhaltsverzeichnis

Aufgabe Ü4.3: *Klassifizierung auf Basis eines einzigen Merkmals*

Aufgabenstellung

In beliebiger Orientierung liegende und einander nicht überlappende gleichseitige Dreiecke und Quadrate mit gleicher Fläche sollen anhand ihres Formfaktors unterschieden werden. Welche Trennschwelle sollte für die Klassifikation gewählt werden?

Aufgabe Ü4.3

Schritt #1: Lösung

Teilschritt #1: Flächenformeln (sei a die Seitenlänge eines gleichseitigen Dreiecks und b – die Seitenlänge eines Quadrats)

Gleichseitiges Dreieck [Satz des Heron](#)	$F_\Delta = \sqrt{\frac{3\cdot a}{2}\cdot\left(\frac{3\cdot a}{2}-a\right)^3} = \sqrt{\frac{3\cdot a}{2^4}\cdot a^3} = \sqrt{3}\cdot\left(\frac{a}{2}\right)^2 = \frac{\sqrt{3}}{4}\cdot a^2$
Quadrat	$F_\square = b^2$

Teilschritt #2: Umfang

Gleichseitiges Dreieck	$U_\Delta = 3\cdot a$
Quadrat	$U_\square = 4\cdot b$

Teilschritt #3: Voraussetzung

$$F_\Delta = F_\square \Longrightarrow \frac{\sqrt{3}}{4}\cdot a^2 = b^2$$

Hausaufgabe: wozu diese Voraussetzung?

Teilschritt #4: Formfaktor

Gleichseitiges Dreieck	$S_\Delta = \frac{{U_\Delta}^2}{4\cdot\pi\cdot F_\Delta} = \frac{(3\cdot a)^2}{4\cdot\pi\cdot\frac{\sqrt{3}}{4}\cdot a^2} = \frac{3\cdot\sqrt{3}}{\pi} \approx 1.65399$
Quadrat	$S_\square = \frac{{U_\square}^2}{4\cdot\pi\cdot F_\square} = \frac{(4\cdot b)^2}{4\cdot\pi\cdot b^2} = \frac{4}{\pi} \approx 1.27324$

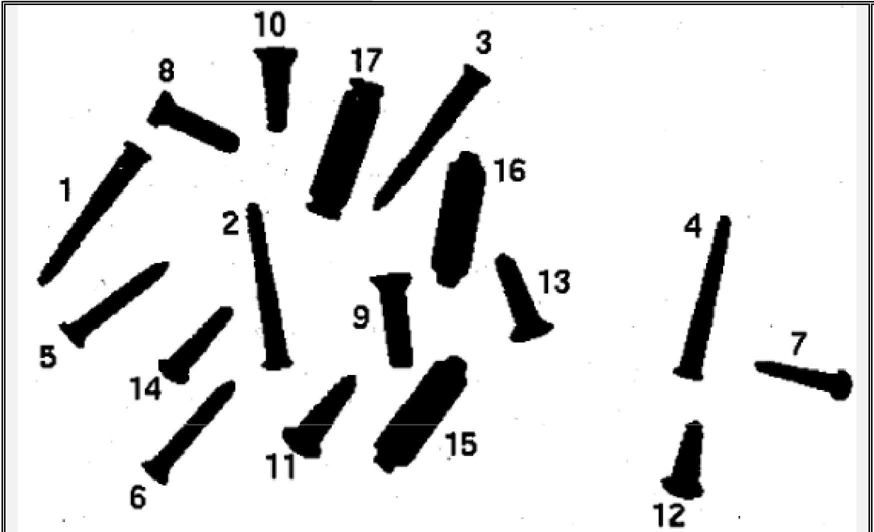
Teilschritt #5: Schwellenwert zwecks optimaler Trennschärfe bei gestörten Daten

$$\frac{S_\Delta + S_\square}{2} \approx 1.4636$$

Hausaufgabe: warum genau arithmetisches Mittel?

Aufgabe Ü6.1: *Abstandsklassifikator - Euklid*

Aufgabenstellung



Bedeutung der Merkmale:	$x1$:	Segmentfläche
	$x2$:	Konturlänge
Bedeutung der Klassen:	$k1$:	schlanke Holzschrauben
	$k2$:	restliche Teile

(a) Stellen Sie die Objekte im zweidimensionalen Merkmalsraum dar.

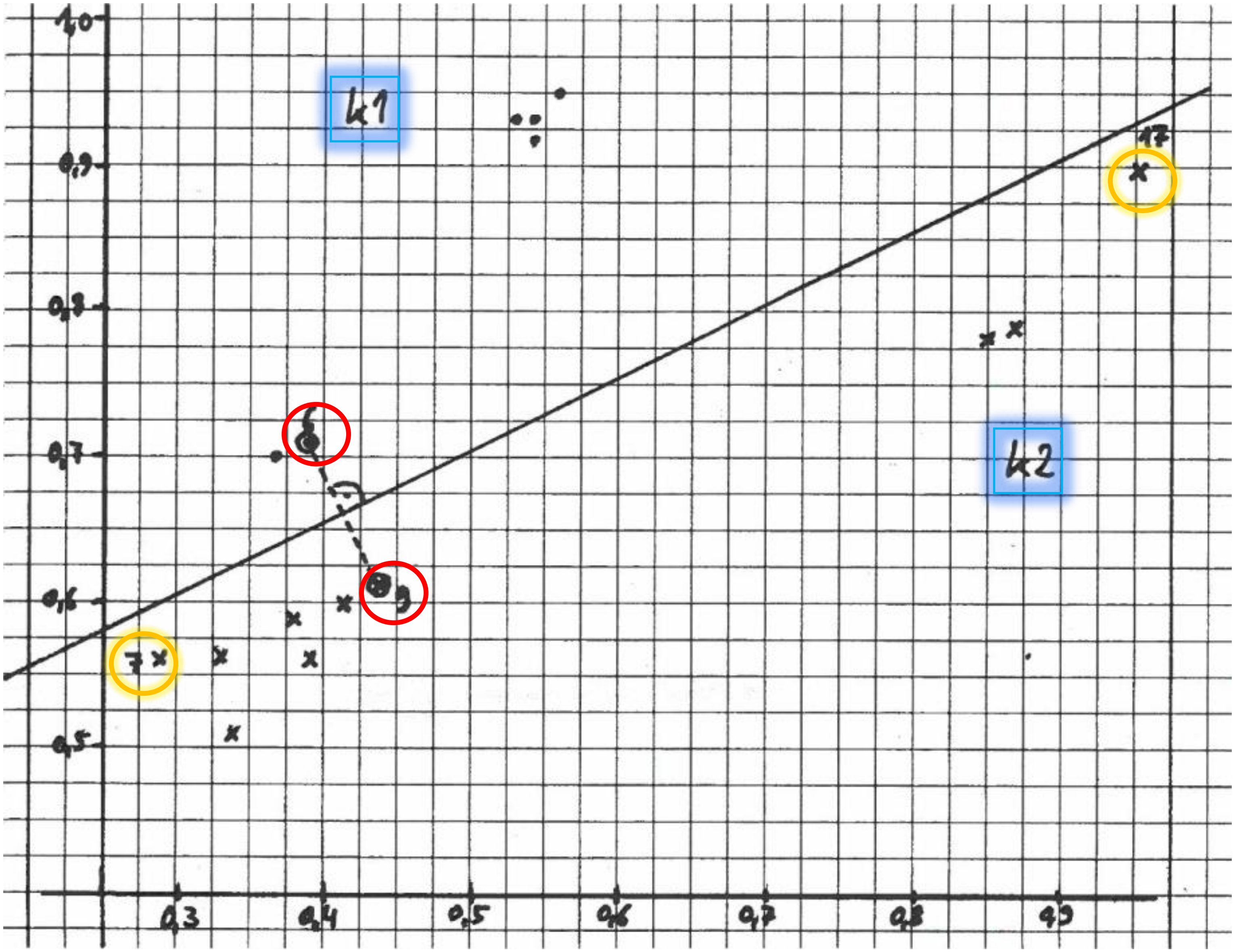
(b) Bestimmen Sie je einen Repräsentanten der Klassen $k1$ und $k2$, so dass der zugehörige Abstandsklassifikator (Objekt wird der Klasse zugeordnet, deren Repräsentant den kleineren euklidischen Abstand im Merkmalsraum hat) die beiden Klassen korrekt trennt. Überprüfen Sie die korrekte Zuordnung an mindestens zwei "Grenzfällen" (Objekte nahe der Trennlinie).

(c) Tragen Sie die vom Klassifikator aus (b) induzierte Trennlinie in das Diagramm aus (a) ein und bestimmen Sie ihre Geradengleichung.

Aufgabe Ü6.1

Schritt #1: Lösung

(a)



Aufgabe Ü6.1

(b)

Teilschritt #1: Repräsentant für k_1

$$Objekt_{\#6} = (0.39; 0.71)$$

Teilschritt #2: Repräsentant für k_2

$$Objekt_{\#9} = (0.44; 0.61)$$

Anmerkung: wie komme ich auf diese und nicht mal auf andere Objekte? ☺

Herkömmliche Methode: Nächste-Nachbarn-Algorithmen [**Dirichlet-Zerlegung** als Sonderfall]

P.S. Auch andere Objekte könnten ebenso gut die Klassen repräsentieren

Mathe-Kurznotiz ☺: man darf hier direkt mit Quadraten von Distanzen ruhig operieren – bei positiven Werten ist ja die Parabel $f(x) = x^2$ steigend

Teilschritt #3: QualitätsKontrolle ☺ - Objekte nahe der Trennlinie

$Objekt_{\#7} = (0.29; 0.56)$ – **korrekte** Zuordnung weil:	
k_1	$d^2_{Euklid}(Objekt_{\#6}; Objekt_{\#7}) = (0.39 - 0.29)^2 + (0.71 - 0.56)^2 = 0.0325$
k_2	$d^2_{Euklid}(Objekt_{\#9}; Objekt_{\#7}) = (0.44 - 0.29)^2 + (0.61 - 0.56)^2 = 0.025$
$Objekt_{\#17} = (0.95; 0.90)$ – **korrekte** Zuordnung weil:	
k_1	$d^2_{Euklid}(Objekt_{\#6}; Objekt_{\#17}) = (0.39 - 0.95)^2 + (0.71 - 0.90)^2 = 0.3497$
k_2	$d^2_{Euklid}(Objekt_{\#9}; Objekt_{\#17}) = (0.44 - 0.95)^2 + (0.61 - 0.90)^2 = 0.3442$

(c)

Teilschritt #1: Trennlinie

Hausaufgabe für Mathe Fans ☺: beweist es oder erfahrt es bei Daniel Jung ☺

Mittelsenkrechte zur Verbindungsstrecke zwischen gewählten Repräsentanten von k_1 und k_2

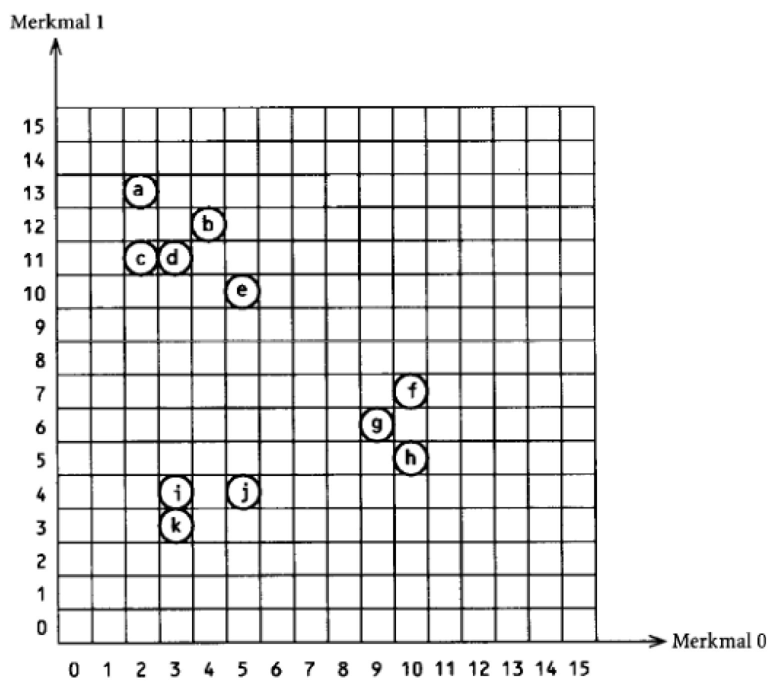
Teilschritt #2: Bestimmung der Geradengleichung

Mittelpunkt der Verbindungsstrecke	$\vec{p} = \frac{1}{2} \cdot (\overrightarrow{p_6} + \overrightarrow{p_9}) = (0.415; 0.66)$
Normalenvektor der Geraden	$\vec{n} = \overrightarrow{p_9} - \overrightarrow{p_6} = (0.05; -0.1)$
Geradengleichung	$\vec{n} \cdot (\vec{x} - \vec{p}) = 0 \Rightarrow \cdots \Rightarrow y = \frac{1}{2} \cdot x + 0.4525$

Aufgabe U19: *Abstandsklassifikator - Mahalanobi*

Aufgabenstellung

Gegeben sind die Objekte *a*–*k* in einem zweidimensionalen Merkmalsraum:



Die Objekte *a*–*e* sollen eine Lernstichprobe für eine Klasse k_0 auf der Grundlage des Mahalanobis-Klassifikators bilden. Die Zurückweisungsschwelle d_0 sei $\vec{\sigma}_0^T \Sigma_0^{-1} \vec{\sigma}_0$.
In der Anwendungsphase des Klassifikators sollen 2 Objekte $p = (5; 10)^T$ und $q = (6; 9)^T$ klassifiziert werden. Gehören sie zu k_0 ?

Aufgabe U19

Schritt #1: Lösung

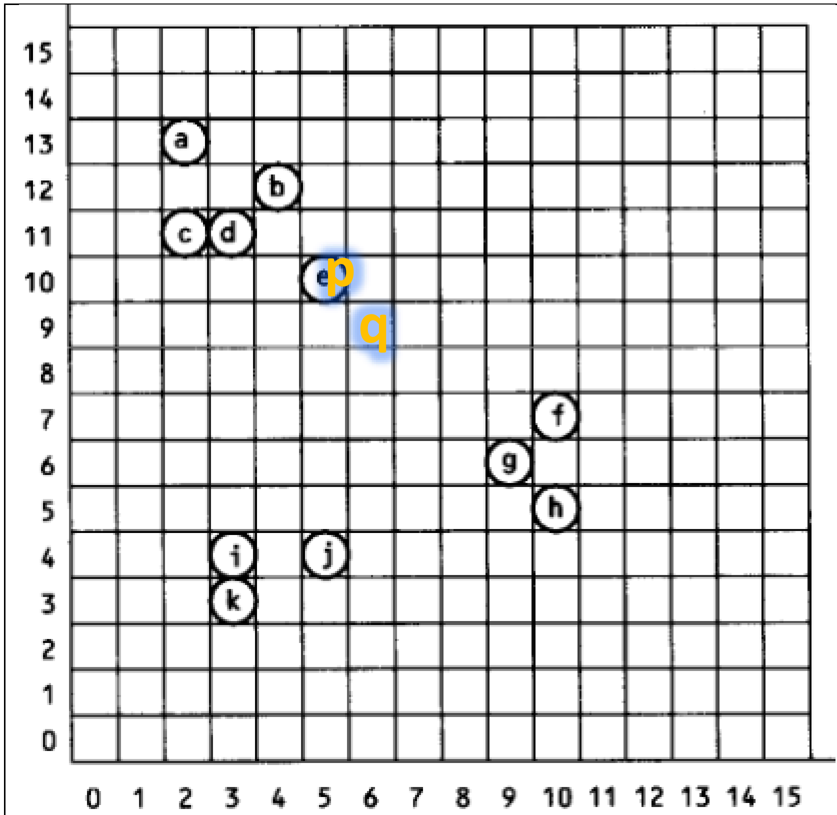
Teilschritt #1: Mahalanobis-Distanz

Interessante Nebenbeobachtungen für Mathe Fans ☺:

Geometrie: ohne Σ_0^{-1} dazwischen hätte man Euklidische Distanz gehabt – ein Kreis, zentriert am Schwerpunkt, mit dem Zeug Σ_0^{-1} bekommt man aber eine **Ellipse**

Algebra: die Kovarianzmatrix Σ_0 liefert allerschönste Eigenschaften:

- durch ihre Symmetrie und Gramm-Schmidtsches Ortogonalisierungsverfahren bekommt man eine Orthonormalbasis von ihren Eigenvektoren (checkt den Spektralsatz) – so kriegt man die $\Sigma_0 = Q\Lambda Q^T$ Darstellung, wobei Λ eine **Diagonalmatrix** bestehend aus ihren **Eigenwerten** und Q – eine **orthogonale Matrix** bestehend aus ihren **Eigenvektoren** ist,
- ihre positiv semidefinite Natur schließt die Negativität von Eigenwerten komplett aus.



Hausaufgabe für Mathe Fans ☺: beweist dass Inverse einer symmetrischen Matrix symmetrisch bleibt

$$d^2_{Mahalanobi} = (\vec{x} - \vec{\mu_0})^T \cdot \Sigma_0^{-1} \cdot (\vec{x} - \vec{\mu_0})$$

Mittelwertvektor der Klasse k_0:

$$\vec{\mu_0} = \frac{1}{|\{a,b,c,d,e\}|} \cdot (a+b+c+d+e)^T = \frac{1}{5} \cdot (16;57)^T$$

Kovarianzmatrix der Klasse k_0:

$$\Sigma_0 = \frac{1}{5-1} \cdot \sum_{j=0}^{5} (\vec{x_j} - \vec{\mu_0}) \cdot (\vec{x_j} - \vec{\mu_0})^T = \cdots = \begin{bmatrix} 1.7 & -0.85 \\ -0.85 & 1.3 \end{bmatrix}$$

Wobei:

$$\vec{x} = (a,b,c,d,e)$$

Inverse zur Kovarianzmatrix:

$$\Sigma_0^{-1} = \frac{1}{det(\Sigma_0)} \begin{bmatrix} \sigma_{11} & -\sigma_{01} \\ -\sigma_{10} & \sigma_{00} \end{bmatrix} = \cdots \approx \begin{bmatrix} 0.874 & 0.571 \\ 0.571 & 1.143 \end{bmatrix}$$

Aufgabe U19

Teilschritt #2: Streuungsvektor $\overrightarrow{\boldsymbol{\sigma_0}}$

$$\overrightarrow{\boldsymbol{\sigma_0}} = \left(\sqrt{\boldsymbol{\sigma_{00}}}; \sqrt{\boldsymbol{\sigma_{11}}}\right)^{\boldsymbol{T}} = \cdots \approx (\mathbf{1.304}; \mathbf{1.140})^{\boldsymbol{T}}$$

Teilschritt #3: Züruckweisungsschwelle

$$(\boldsymbol{d_0})^2 = \overrightarrow{\boldsymbol{\sigma_0}}^{\boldsymbol{T}} \cdot \boldsymbol{\Sigma_0^{-1}} \cdot \overrightarrow{\boldsymbol{\sigma_0}} = \cdots \approx \mathbf{4.67}$$

Teilschritt #4: Mahalanobis-Distanz für $\vec{\boldsymbol{p}}$ und $\vec{\boldsymbol{q}}$

$$\boldsymbol{d^2_{Mahalanobi}}(\vec{\boldsymbol{p}}, \overrightarrow{\boldsymbol{\mu_0}}) = \cdots \approx \mathbf{2.195} < (\boldsymbol{d_0})^2 \Longrightarrow \vec{\boldsymbol{p}} \text{ wird zu } \boldsymbol{k_0} \text{ klassifiziert}$$

$$\boldsymbol{d^2_{Mahalanobi}}(\vec{\boldsymbol{q}}, \overrightarrow{\boldsymbol{\mu_0}}) = \cdots \approx \mathbf{5.762} > (\boldsymbol{d_0})^2 \Longrightarrow \vec{\boldsymbol{q}} \text{ wird nicht zu } \boldsymbol{k_0} \text{ klassifiziert}$$