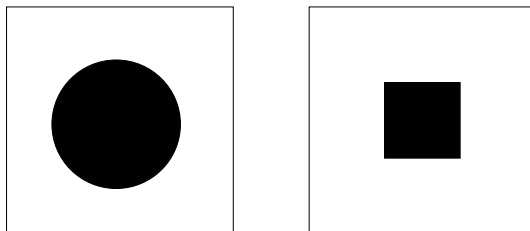


Kapitel 6

Optimale Merkmalsselektion

Einfaches Erkennungsbeispiel mit zwei Objekten (Kreis und Quadrat)



Es wird angenommen, dass die Bilder mit 512×512 Bildpunkte abgetastet werden.

Niemand käme auf die Idee hier 512×512 Bildpunkte als Merkmale der Objekte zu verwenden!! (1 Merkmal würde reichen: Fläche)

Merkmalsselektion mit linearen Transformationen

Die Komplexität eines Klassifikatorentwurfs wächst mit der Dimension N des Merkmalsraums. Ziel einer Merkmalsselektion ist die Auswahl eines geeigneten Unterraumes. Die selektierten Merkmale müssen eine hohe Relevanz für die Charakterisierung der Klassen besitzen, aber zugleich eine hohe Diskriminierfähigkeit zwischen den Klassen garantieren. Sie müssen demnach innerhalb einer Klasse wenig variieren (Intraklassenabstand), aber gleichzeitig große Abstände zwischen den Klassen (Interklassenabstand) garantieren.

Es ist i.allg. wenig sinnvoll die Pixel eines Bildes direkt als Merkmale zu verwenden ($N=512^2=2^{18}=0,25$ Mio. Pixel). I.allg. gibt es eine hohe Redundanz in den Bildern durch starke Korrelation zwischen den Pixeln.

Es ist zudem wenig hilfreich einen Merkmalsraum durch Hinzunahme neuer Merkmale weiter zu vergrößern, wenn die neuen Merkmale stark korreliert sind mit den bereits vorhandenen.

Idee: Transformation der Originalbilder in einen neuen Merkmalsraum (Verschieben und Drehung des Koordinatensystems (unitäre Transformation)). Dabei Reduktion auf wenige Merkmale und gleichzeitige Informationsverdichtung.

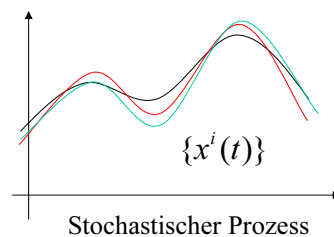
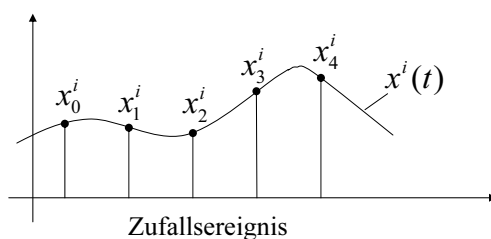
Charakterisierung von Zufallsereignissen in Vektorräumen

Ein *Zufallsereignis* \mathbf{x}^i ist ein Element des Vektorraumes \mathbb{X} . Wobei im diskreten Fall das Elementarereignis aus einer geordneten Menge von Zahlenwerten

$$\mathbf{x}^i := \{x_0^i, x_1^i, x_2^i, \dots, x_{N-1}^i\}$$

oder auch im kontinuierlichen Fall aus einer Zeit- oder Ortsfunktionen $x^i(t)$ besteht.

Ein *stochastischer Prozess* \mathbf{x} besteht aus einer Menge von Ereignissen $\mathbf{x} := \{\mathbf{x}^j\}$.



Statistische Kenngrößen eines Prozesses

Erwartungswert: $\mu_x = \bar{x} = E\{\mathbf{x}\} \approx \frac{1}{N} \sum_{i=0}^{N-1} \mathbf{x}^i$

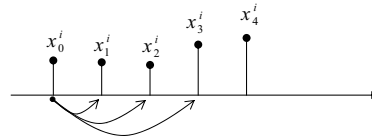
Autokorrelation: $\mathbf{R}_{xx} = E\{\underbrace{\mathbf{xx}^T}_{\substack{\text{dyad.} \\ \text{Produkt!}}}\} = \{E(x_i x_j)\} \approx \frac{1}{N} \sum_{i=0}^{N-1} \begin{bmatrix} x_0^i \cdot x_0^i & x_0^i \cdot x_1^i & x_0^i \cdot x_2^i \\ x_1^i \cdot x_0^i & x_1^i \cdot x_1^i & x_1^i \cdot x_2^i \\ x_2^i \cdot x_0^i & x_2^i \cdot x_1^i & x_2^i \cdot x_2^i \end{bmatrix}$

Autokovarianz: $\mathbf{C}_{xx} = E\{(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T\} = \mathbf{R}_{xx} - \bar{\mathbf{xx}}^T$

Kreuzkorrelation: $\mathbf{R}_{xy} = E\{\mathbf{xy}^T\}$

Kreuzkovarianz: $\mathbf{C}_{xy} = E\{(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{y} - \bar{\mathbf{y}})^T\} = \mathbf{R}_{xy} - \bar{\mathbf{xy}}^T$

Die Elemente der Korrelationsmatrix beschreiben die Korrelation zwischen den einzelnen Vektorelementen $\{x_0, x_1\}$, $\{x_0, x_2\}$... in zeitlicher/örtlicher Richtung mit wachsendem Abstand zwischen den Elementen:



Bei *linearen* Transformationen bleiben Gaußverteilungen erhalten

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N \det(\mathbf{C}_{xx})}} e^{-\frac{1}{2}(\mathbf{x} - \mu_x)^T \mathbf{C}_{xx}^{-1} (\mathbf{x} - \mu_x)}$$

aus $\mathbf{y} = \mathbf{Ax}$ folgt für $p(\mathbf{y})$ wieder eine Normalverteilung mit:

$$\mu_y = \mathbf{A}\mu_x$$

und:

$$\mathbf{C}_{yy} = E\{(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})^T\} = \mathbf{A}E\{(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T\}\mathbf{A}^T$$

$$\Rightarrow \mathbf{C}_{yy} = \mathbf{A}\mathbf{C}_{xx}\mathbf{A}^T$$

Berechnung der AKF aus der Autokorrelationsmatrix

Die Werte der linearen (zyklischen) AKF ergeben sich aus den Diagonalsummen der (periodisch fortgesetzten) Autokorrelationsmatrix:

lineare AKF:

$$\begin{array}{ccc} x_0 & x_1 & x_2 \\ \hline x_0^2 + x_1^2 + x_2^2 & x_0 x_1 + x_1 x_2 & x_0 x_2 \end{array}$$

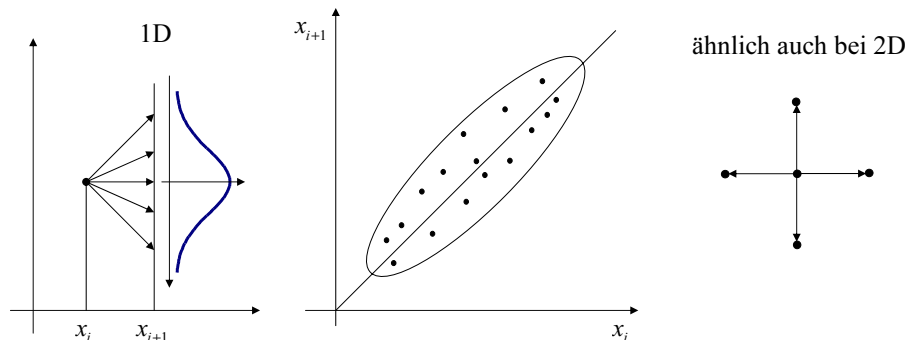
Autokorrelationsmatrix:

	x_0	x_1	x_2
x_0	$x_0 x_0$	$x_0 x_1$	$x_0 x_2$
x_1	$x_1 x_0$	$x_1 x_1$	$x_1 x_2$
x_2	$x_2 x_0$	$x_2 x_1$	$x_2 x_2$

Dekorrelation benachbarter Signal- oder Pixelwerte im Vektorraum

Gegeben ein Signal- oder Pixelwert; wie groß ist die W., daß die benachbarte Signalamplitude ähnliche Werte annimmt?

I.a. hohe Korrelation in die Nachbarschaft! (1. Winkelhalbierende im Vektorraum)



Einfaches Beispiel für den Übergang in einen neuen Merkmalsraum mit Hilfe einer orthogonalen Transformation

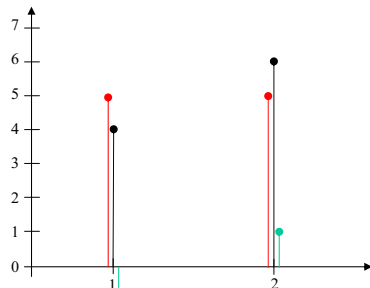
Gegeben sei ein Signal mit zwei Abtastwerten

$$\mathbf{x} = \begin{bmatrix} 4 \\ 6 \end{bmatrix} = 4 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 6 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Originalraum

$$= 5\sqrt{2} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \sqrt{2} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Merkmalsraum



Selektiert man nur die erste Komponente (Unterraum) im Originalraum, so erhält man eine Approximationsgüte von (ein Abtastwert weglassen fällt auf!):

$$\frac{\| \begin{bmatrix} 4 & 0 \end{bmatrix} \|}{\| \begin{bmatrix} 4 & 6 \end{bmatrix} \|} = \frac{4}{7,21} = \boxed{55\%}$$

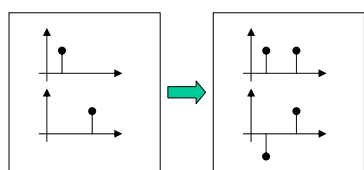
Im neuen Merkmalsraum hingegen:

$$\frac{\| \begin{bmatrix} 5\sqrt{2} & 0 \end{bmatrix} \|}{\| \begin{bmatrix} 5\sqrt{2} & \sqrt{2} \end{bmatrix} \|} = \frac{7,07}{7,21} = \boxed{98\%}$$

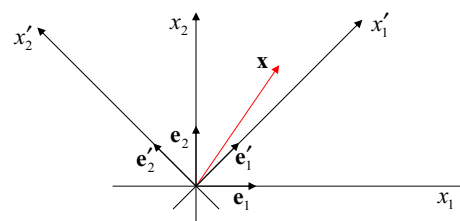
Außerdem sind die neuen Werte weniger korreliert!

— Gleichanteil
— Wechselanteil

Darstellung im Vektorraum durch Drehung des Koordinatensystems mit orthogonaler Transformation



Originalraum Merkmalsraum



$$\boxed{\mathbf{x}' = \mathbf{A}^T \mathbf{x}}$$

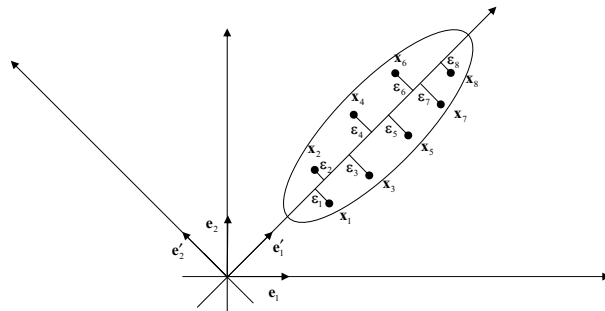
mit: $\mathbf{A}^T = \frac{1}{\sqrt{2}} \begin{bmatrix} +1 & +1 \\ -1 & +1 \end{bmatrix}$ und: $(\mathbf{A}^T)^{-1} = (\mathbf{A}^T)^T = \mathbf{A} = \frac{1}{\sqrt{2}} \begin{bmatrix} +1 & -1 \\ +1 & +1 \end{bmatrix}$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = x'_1 \mathbf{e}'_1 + x'_2 \mathbf{e}'_2 = x'_1 \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x'_2 \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \underbrace{\frac{(x_1 + x_2)}{2}}_{\text{Gleichanteil}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \underbrace{\frac{(x_2 - x_1)}{2}}_{\text{Wechselanteil}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

mit: $\mathbf{e}'_1 = \mathbf{A} \mathbf{e}_1$ und $\mathbf{e}'_2 = \mathbf{A} \mathbf{e}_2$

Optimale Merkmalsselektion mit unitären Transformationen

(Karhunen-Loeve oder Hauptachsentransformation)



$$\{e_i\} \xrightarrow{A^T} \{e'_i\}$$

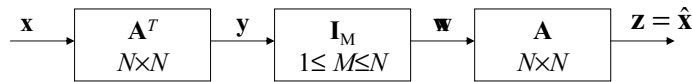
$$\boxed{x' = A^T x}$$

Ein einziges Vektorelement x kann, wenn das dazugehörige Basissystem frei gewählt werden kann (und auf Sende- und Empfangsseite bekannt ist) durch einen skalaren Wert charakterisiert werden, wenn der erste Basisvektor e'_1 in Richtung x gewählt wird (Element kommt vor oder nicht):

$$x = \alpha \frac{x}{\|x\|} + 0 \cdot e'_2 + 0 \cdot e'_3 + \dots$$

Es geht jedoch i.allg. darum, für ein gesamtes *Ensemble von Vektoren* eine optimale Transformation in ein geeignetes Koordinatensystem zu finden, so dass im Mittel die Elemente des Ensembles mit möglichst wenig Koeffizienten charakterisiert werden können.

Wir beginnen mit der Bestimmung des ersten neuen Basisvektors e'_1 , welcher so gewählt wird, dass der *Approximationsfehler* für das Ensemble von n Vektoren *minimal* wird, oder die *gesuchte Raumrichtung*, die eine *maximale Information* des Ensembles repräsentiert. Nach dem Projektionssatz erhält man den kleinsten Fehler bei Projektion auf den Unterraum, welcher repräsentiert wird durch e'_1 ; gesucht ist lediglich die richtige Raumrichtung. Ausgehend von einem Gütekriterium, wird eine optimale Lösung gesucht.



$$\mathbf{w} = \mathbf{I}_M \mathbf{y} = y_1 \mathbf{e}'_1 + y_2 \mathbf{e}'_2 + \dots + y_M \mathbf{e}'_M$$

die Bestapproximation muss unabhängig vom M sein!

Ausgehend von einem quadratischen Gütekriterium ergibt sich:

$$\begin{aligned} J &= \frac{1}{n} \{ \|\boldsymbol{\varepsilon}_1\|^2 + \|\boldsymbol{\varepsilon}_2\|^2 + \dots + \|\boldsymbol{\varepsilon}_n\|^2 \} \\ &= \frac{1}{n} \{ \|\mathbf{x}_1 - \langle \mathbf{x}_1, \mathbf{e}'_1 \rangle \mathbf{e}'_1\|^2 + \|\mathbf{x}_2 - \langle \mathbf{x}_2, \mathbf{e}'_1 \rangle \mathbf{e}'_1\|^2 + \dots \} \\ &= E \{ \|\mathbf{x} - \underbrace{\langle \mathbf{x}, \mathbf{e}'_1 \rangle \mathbf{e}'_1}_{\substack{\mathbf{Px} \\ \text{Projektion auf} \\ S(\mathbf{e}'_1)}}\|^2 \} \end{aligned}$$

für das Ensemble: $\mathbf{x} := \{\mathbf{x}_i\} \quad i = 1, 2, \dots, n$

ZR: Bei einer orthogonalen Projektion auf einen Unterraum \mathbf{Px} gilt gemäß dem Satz von Pythagoras:

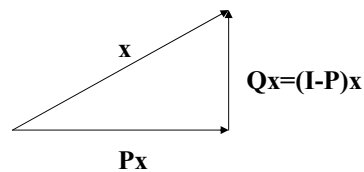
$$\|\mathbf{x}\|^2 = \|\mathbf{Px}\|^2 + \|\underbrace{(\mathbf{I} - \mathbf{P})\mathbf{x}}_{\mathbf{Q}}\|^2$$

$$\Rightarrow \|\mathbf{x} - \mathbf{Px}\|^2 = \|\mathbf{x}\|^2 - \|\mathbf{Px}\|^2$$

$$\text{mit: } \langle \mathbf{x}, \mathbf{e}'_1 \rangle \mathbf{e}'_1 = \underbrace{(\mathbf{e}'_1 \mathbf{e}'_1{}^T)}_{\mathbf{P}} \mathbf{x}$$

bei reellen Vektoren gilt:

$$\langle \mathbf{a}, \mathbf{b} \rangle = \langle \mathbf{b}, \mathbf{a} \rangle$$



und deshalb:

$$\begin{aligned}
 & \langle \langle \mathbf{x}, \mathbf{e}'_1 \rangle \mathbf{e}'_1, \langle \mathbf{x}, \mathbf{e}'_1 \rangle \mathbf{e}'_1 \rangle \\
 & = \langle \mathbf{x}, \mathbf{e}'_1 \rangle^2 \underbrace{\langle \mathbf{e}'_1, \mathbf{e}'_1 \rangle}_{=1} \\
 J & = E \left\{ \|\mathbf{x}\|^2 - \|\langle \mathbf{x}, \mathbf{e}'_1 \rangle \mathbf{e}'_1\|^2 \right\} \\
 & = E \left\{ \|\mathbf{x}\|^2 - \langle \mathbf{x}, \mathbf{e}'_1 \rangle^2 \right\} \quad (\text{Maximierung der Quadrate der FK}) \\
 & = E \left\{ \|\mathbf{x}\|^2 - \langle \mathbf{e}'_1, \mathbf{x} \rangle \langle \mathbf{x}, \mathbf{e}'_1 \rangle \right\}
 \end{aligned}$$

Nützliche Formeln:

$$\langle \mathbf{a}, \mathbf{b} \rangle \langle \mathbf{c}, \mathbf{d} \rangle = \mathbf{a}^T \overbrace{(\mathbf{b}\mathbf{c}^T)}^{\text{dyad. Produkt}} \mathbf{d}$$

$$\mathbf{a} \langle \mathbf{b}, \mathbf{c} \rangle = (\mathbf{a}\mathbf{b}^T) \mathbf{c}$$

$$\langle \mathbf{a}, \mathbf{b} \rangle = \text{Spur}(\mathbf{a}\mathbf{b}^T)$$

Das Innenprodukt kann über die Spur des Aussenproduktes berechnet werden!

$$\langle \mathbf{A}, \mathbf{B} \rangle = \text{Spur}(\mathbf{A} \underbrace{\mathbf{B}^*}_{\mathbf{B} \text{ adjungiert}})$$

bei Bildmatrizen

und deshalb:

$$\begin{aligned}
 J & = E \left\{ \|\mathbf{x}\|^2 - \mathbf{e}'_1{}^T (\mathbf{x}\mathbf{x}^T) \mathbf{e}'_1 \right\} \\
 & = E \left\{ \underbrace{\|\mathbf{x}\|^2}_{\sigma_x^2 \text{ Varianz von } x} \right\} - \mathbf{e}'_1{}^T \underbrace{E \{ \mathbf{x}\mathbf{x}^T \}}_{\mathbf{R}_{xx} \text{ Auto-korrelationsmatrix}} \mathbf{e}'_1 \\
 \Rightarrow & \boxed{J = \text{Spur}(\mathbf{R}_{xx}) - \mathbf{e}'_1{}^T \mathbf{R}_{xx} \mathbf{e}'_1 = \min_{\mathbf{e}'_1} !}
 \end{aligned}$$

Als Nebenbedingung geht ein, dass es sich bei dem neuen Basisvektor um einen Einheitsvektor handelt:

$$\|\mathbf{e}'_1\|^2 = \langle \mathbf{e}'_1, \mathbf{e}'_1 \rangle = 1$$