7 Polynomklassifikator

Ziel: Optimierung der Unterscheidungsfunktion $\vec{d}(\vec{c}_i)$ bezüglich der Zielvektoren $\vec{y}_i$ mit dem Quadratmittelansatz

$$S^2 = E\{(\vec{y}(\vec{c}) - \vec{d}(\vec{c}))^2\} \to \min$$

ohne direkte Schätzung von Dichten

- in Abschnitt 6.4 beliebige Funktionen für die Optimierung
  $\to \vec{d}(\vec{c}) = (p(\omega_1 \mid \vec{c}, \ldots)$
- jetzt Beschränkung auf eine Funktionenklasse
  $\to$ Problem der Parameterschätzung
- wünschenswert sind Funktionenklassen, die jede beliebige Funktion annähern können $\to$ universeller Approximator

7.1 Der Ansatz

Polynome sind ein universeller Approximator wegen des Satzes von Weierstraß:

Jede stetige Funktion kann beliebig genau durch Polynome approximiert werden, sofern der Polynomgrad groß genug gewählt wird.

Modellierung der Komponenten der Unterscheidungsfunktion $d_i(\vec{c})$ wie folgt:

$$
\begin{aligned}
d_i(\vec{c}) &= a_{i0} + a_{i1}c_1 + a_{i2}c_2 + \ldots + a_{iN}c_N + \\
&\quad a_{i,(N+1)}c_1^2 + a_{i,(N+2)}c_1c_2 + a_{i,(N+3)}c_1c_3 + \ldots + \\
&\quad a_{i,(N+N(N+1)/2+1)}c_1^3 + a_{i,(N+N(N+1)/2+2)}c_1^2c_2 + a_{i,(N+N(N+1)/2+3)}c_1^2c_3 + \ldots \\
&= \vec{a}_i^T \vec{x}(\vec{c})
\end{aligned}
$$

mit $\vec{x}(\vec{c}) = (1, c_1, \ldots, c_N, c_1^2, c_1c_2, \ldots, c_N^2, c_1^3, \ldots)^T$

und $dim(\vec{c}) = N$

- Unterscheidungsfunktion wird über Polynome in den Komponenten des Merkmalsvektors approximiert
- zu optimierende Parameter: $\vec{a}_i$

für die Unterscheidungsfunktion $\vec{d}(\vec{c})$ gilt damit:

$$\vec{d}(\vec{c}) = \begin{pmatrix} d_1(\vec{c}) \\ \vdots \\ d_K(\vec{c}) \end{pmatrix} = \begin{pmatrix} \vec{a}_1^T \vec{x}(\vec{c}) \\ \vdots \\ \vec{a}_K^T \vec{x}(\vec{c}) \end{pmatrix} = \underline{A}^T \vec{x}(\vec{c})$$

$$\text{mit} \quad \underline{A} = (\vec{a}_1, \ldots, \vec{a}_K)$$

Vor dem Training eines Polynomklassifikators ist festzulegen:

- Polynomgrad G
- Polynomstruktur: z.B. welche quadratischen Terme $c_i c_k$ verwendet werden

meist vollständige Polynome bis zum Grad G

Beachte: Grad nicht zu groß, da sich bei N-dimensionalem Merkmalsvektor folgende Dimension für $\vec{x}(\vec{c})$ ergibt:

$$M = \binom{N+G}{G} = dim(\vec{x}(\vec{c}))$$

z.B.

			$K = 30$
$N = 10$	$G = 2$	$M = 66$	1980
$N = 30$	$G = 2$	$M = 496$	14880
$N = 30$	$G = 3$	$M = 1891$	56730
$N = 60$	$G = 3$	$M = 39711$	1191330

Bemerkung: im Prinzip kann man für die $x_i(\vec{c})$ nicht nur Monome über den c_i zulassen, sondern beliebige Funktionen verwenden
$\rightarrow$ Unterscheidungsfunktion ist Linearkombination von beliebigen Funktionen
$\rightarrow$ für die Optimierung von $\underline{A}$ spielt dies keine Rolle.

7.2 Lösung des Minimierungsproblems

nun mit Hilfe der Variationsrechnung die optimale Koeffizientenmatrix bestimmen:

Optimierungskriterium :

$$S^2 = E\left\{\left|\vec{y}(\vec{c}) - \vec{d}(\vec{c})\right|^2\right\} = E\left\{\left|\vec{y}(\vec{c}) - \underline{A}^T\vec{x}(\vec{c})\right|^2\right\} = S^2(\underline{A}) \to \min_{\underline{A}}$$

(im weiteren statt $\vec{x}(\vec{c})$ und $\vec{y}(\vec{c})$ nur noch $\vec{x}$ bzw. $\vec{y}$)

sei $\underline{A}$ die optimale Matrix,
dann verschlechter jede Abweichung $\delta\underline{A} \neq 0$ das Optimierungskriterium:

$$\forall \delta\underline{A} \neq 0 : \quad S^2(\underline{A} + \delta\underline{A}) \geq S^2(\underline{A}) \tag{7.1}$$

Einschub

- für Vektoren gleicher Dimension gilt: $\vec{a}^T\vec{b} =$ spur $\left(\vec{a}\vec{b}^T\right)$
- für quadratische Matrizen gleicher Dimension gilt: spur$(\underline{AB^T}) =$ spur$(\underline{BA^T})$

Somit ergibt sich für die rechte Seite der vorangehenden Ungleichung:

$$
\begin{aligned}
S^2(\underline{A}) &= E\left\{\left|\vec{y} - \underline{A}^T\vec{x}\right|^2\right\} \\
&= E\left\{\left[\vec{y} - \underline{A}^T\vec{x}\right]^T \left[\vec{y} - \underline{A}^T\vec{x}\right]\right\} \\
&= E\left\{\text{spur}\left[\left[\vec{y} - \underline{A}^T\vec{x}\right] \left[\vec{y} - \underline{A}^T\vec{x}\right]^T\right]\right\} \\
&= \text{spur}\left[E\left\{\vec{y}\vec{y}^T\right\}\right] - \text{spur}\left[E\left\{\vec{y}\vec{x}^T\right\}\underline{A}\right] - \text{spur}\left[\underline{A}^T E\left\{\vec{x}\vec{y}^T\right\}\right] + \text{spur}\left[A^T E\left\{\vec{x}\vec{x}^T\right\}\underline{A}\right] \\
&= \text{spur}\left[E\left\{\vec{y}\vec{y}^T\right\}\right] - 2\cdot\text{spur}\left[\underline{A}^T E\left\{\vec{x}\vec{y}^T\right\}\right] + \text{spur}\left[A^T E\left\{\vec{x}\vec{x}^T\right\}\underline{A}\right]
\end{aligned}
$$

analoge Rechnung für die linke Seite:

$$
\begin{aligned}
S^2(\underline{A} + \delta\underline{A}) = \; & \text{spur}\left[E\left\{\vec{y}\vec{y}^T\right\}\right] - 2\cdot\text{spur}\left[\underline{A}^T E\left\{\vec{x}\vec{y}^T\right\}\right] + \text{spur}\left[\underline{A}^T E\left\{\vec{x}\vec{x}^T\right\}\underline{A}\right] \\
& 2\cdot\text{spur}\left[\delta\underline{A}^T\left(E\left\{\vec{x}\vec{y}^T\right\} - E\left\{\vec{x}\vec{x}^T\right\}\underline{A}\right)\right] \\
& +\text{spur}\left[\delta\underline{A}^T E\left\{\vec{x}\vec{x}^T\right\}\delta\underline{A}\right]
\end{aligned}
$$

Einsetzen in die Ungleichung (7.1) ergibt:

$$
\begin{aligned}
S^2(\underline{A} + \delta\underline{A}) \;&\geq\; S^2(\underline{A}) \quad \Leftrightarrow \\
\text{spur}\left[\delta\underline{A}^T E\left\{\vec{x}\vec{x}^T\right\}\delta\underline{A}\right] - 2\cdot\text{spur}\left[\delta\underline{A}^T\left(E\left\{\vec{x}\vec{y}^T\right\} - E\left\{\vec{x}\vec{x}^T\right\}\underline{A}\right)\right] \;&\geq\; 0
\end{aligned}
$$

- spur $\left[\delta\underline{A}^T E\left\{\vec{x}\vec{x}^T\right\}\delta\underline{A}\right] \geq 0$,
  da $E\left\{\vec{x}\vec{x}^T\right\}$ als Korrelationsmatrix positiv semi-definit
- obige Ungleichung ist dann notwendigerweise für beliebige $\delta\underline{A}$ erfüllt, falls
  spur $\left[\delta\underline{A}^T\left(E\left\{\vec{x}\vec{y}^T\right\} - E\left\{\vec{x}\vec{x}^T\right\}\underline{A}\right)\right] = 0$,
- also wird $\underline{A}$ optimal, wenn gilt

$$
\begin{aligned}
E\left\{\vec{x}\vec{y}^T\right\} - E\left\{\vec{x}\vec{x}^T\right\}\underline{A} \;&=\; \mathbb{O} \quad \Leftrightarrow \\
\underline{A} \;&=\; \left(E\left\{\vec{x}\vec{x}^T\right\}\right)^{-1} E\left\{\vec{x}\vec{y}^T\right\}
\end{aligned}
$$

7.3 Berechnung der Koeffizientenmatrix

Es stellen sich zwei Probleme:

- Wie berechnet man die Erwartungswerte $E\left\{\vec{x}\vec{x}^T\right\}$ und $E\left\{\vec{x}\vec{y}^T\right\}$?
- Was macht man, falls die Inverse von $E\left\{\vec{x}\vec{x}^T\right\}$ nicht existiert, d.h. es existieren linear abhängige Zeilen (Komponenten von $\vec{x}$ sind linear abhängig)

Lösung Problem 1: Erwartungswert und Kovarianzmatrix haben wir schon geschätzt

analog hier aus der Stichprobe mit I klassifizierten Paaren $(\vec{c}_i, \vec{y}(\vec{c}_i))$:

$$\widehat{E}\left\{\vec{x}\vec{x}^T\right\} = \frac{1}{I}\sum_{i=1}^{I} \vec{x}(\vec{c}_i)\ \vec{x}(\vec{c}_i)^T$$

$$\widehat{E}\left\{\vec{x}\vec{y}^T\right\} = \frac{1}{I}\sum_{i=1}^{I} \vec{x}(\vec{c}_i)\ \vec{y}(\vec{c}_i)^T$$

Lösung Problem 2:

sei $\vec{z}^T = \left(\vec{x}^T, \vec{y}^T\right)$

und $\underline{M}_{\vec{z}} = E\left\{\vec{z}\vec{z}^T\right\} = E\left\{\begin{pmatrix} \vec{x} \\ \vec{y} \end{pmatrix} \left(\vec{x}^T \; \vec{y}^T\right)\right\} = \begin{pmatrix} E\left\{\vec{x}\vec{x}^T\right\} & E\left\{\vec{x}\vec{y}^T\right\} \\ E\left\{\vec{y}\vec{x}^T\right\} & E\left\{\vec{y}\vec{y}^T\right\} \end{pmatrix}$

dabei ist

- $E\left\{\vec{x}\vec{x}^T\right\}$ eine $M \times M$ Matrix
- $E\left\{\vec{x}\vec{y}^T\right\}$ eine $M \times K$ Matrix
- $E\left\{\vec{y}\vec{x}^T\right\}$ eine $K \times M$ Matrix
- $E\left\{\vec{y}\vec{y}^T\right\}$ eine $K \times K$ Matrix
- mit $M = dim(\vec{x})$ und $K = dim(\vec{y})$

unter der Annahme, daß die Inverse von $E\left\{\vec{x}\vec{x}^T\right\}$ existiert, definieren wir folgende Matrix:

$$\underline{T} = \begin{pmatrix} \left[E\left\{\vec{x}\vec{x}^T\right\}\right]^{-1} & \mathbb{O} \\ -E\left\{\vec{y}\vec{x}^T\right\}\left[E\left\{\vec{x}\vec{x}^T\right\}\right]^{-1} & \mathbb{I} \end{pmatrix}$$

dabei ist

- $\left[E\left\{\vec{x}\vec{x}^T\right\}\right]^{-1}$ eine $M \times M$, Matrix
- $E\left\{\vec{y}\vec{x}^T\right\}\left[E\left\{\vec{x}\vec{x}^T\right\}\right]^{-1}$ eine $[K \times M] \cdot [M \times M] = K \times M$ Matrix
- $\mathbb{O}$ ist eine $M \times K$ große Nullmatrix
- $\mathbb{I}$ eine $K \times K$ große Identitätsmatrix.

es gilt

$$\underline{T} \cdot \underline{M}_{\vec{z}} = \begin{pmatrix} \mathbb{I} & \left[E\left\{\vec{x}\vec{x}^T\right\}\right]^{-1} E\left\{\vec{x}\vec{y}^T\right\} \\ \mathbb{O} & -E\left\{\vec{y}\vec{x}^T\right\}\left[E\left\{\vec{x}\vec{x}^T\right\}\right]^{-1} E\left\{\vec{x}\vec{y}^T\right\} + E\left\{\vec{y}\vec{y}^T\right\} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbb{I} & \underline{A} \\ \mathbb{O} & E\left\{\Delta\vec{d}\Delta\vec{d}^T\right\} \end{pmatrix}$$

da mit $\Delta\vec{d} := \vec{y} - \underline{A}^T\vec{x}$ gilt:

$$\begin{aligned}
E\left\{\Delta\vec{d}\Delta\vec{d}^T\right\} &= E\left\{\left(\vec{y} - \underline{A}^T\vec{x}\right)\left(\vec{y} - \underline{A}^T\vec{x}\right)^T\right\} \\
&= E\left\{\vec{y}\vec{y}^T\right\} - E\left\{\vec{y}\vec{x}^T\right\}\underline{A} - \underline{A}^T E\left\{\vec{x}\vec{y}^T\right\} + \underline{A}^T \underbrace{E\left\{\vec{x}\vec{x}^T\right\}\underline{A}}_{E\left\{\vec{x}\vec{y}^T\right\}} \\
&= E\left\{\vec{y}\vec{y}^T\right\} - E\left\{\vec{y}\vec{x}^T\right\}\underline{A} - \underline{A}^T E\left\{\vec{x}\vec{y}^T\right\} + \underline{A}^T E\left\{\vec{x}\vec{y}^T\right\} \\
&= E\left\{\vec{y}\vec{y}^T\right\} - E\left\{\vec{y}\vec{x}^T\right\}\left[E\left\{\vec{x}\vec{x}^T\right\}\right]^{-1} E\left\{\vec{x}\vec{y}^T\right\}
\end{aligned}$$

(beachte: $\underline{A} = \left(E\left\{\vec{x}\vec{x}^T\right\}\right)^{-1} E\left\{\vec{x}\vec{y}^T\right\}$)

zusätzlich gilt (siehe Einschub:)

$$S^2(\underline{A}) = E\left\{\Delta\vec{d}^T\Delta\vec{d}\right\} = \text{spur}(E\left\{\Delta\vec{d}\Delta\vec{d}^T\right\})$$

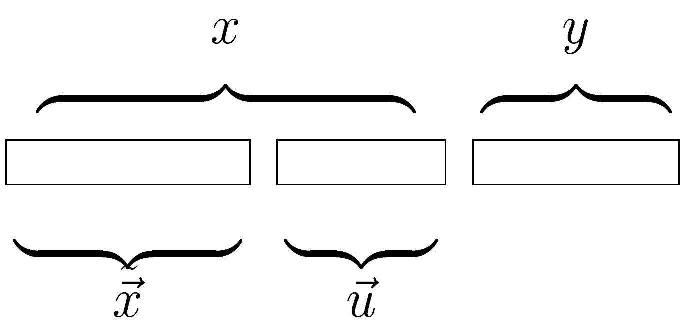
- scheinbar nichts gewonnen, da für $\underline{T}$ immer noch die inverse von $E\left\{\vec{x}\vec{x}^T\right\}$ erforderlich:

$$\begin{pmatrix} \left[E\left\{\vec{x}\vec{x}^T\right\}\right]^{-1} & \mathbb{O} \\ -E\left\{\vec{y}\vec{x}^T\right\}\left[E\left\{\vec{x}\vec{x}^T\right\}\right]^{-1} & \mathbb{I} \end{pmatrix} \begin{pmatrix} E\left\{\vec{x}\vec{x}^T\right\} & E\left\{\vec{x}\vec{y}^T\right\} \\ E\left\{\vec{y}\vec{x}^T\right\} & E\left\{\vec{y}\vec{y}^T\right\} \end{pmatrix} = \begin{pmatrix} \mathbb{I} & \underline{A} \\ \mathbb{O} & E\left\{\Delta\vec{d}\Delta\vec{d}^T\right\} \end{pmatrix}$$

- Linksmultiplikationen von Matrizen entsprechen jedoch elementaren Zeilenumformungen.
- damit reicht es aus, die Matrix $\underline{M}_{\vec{z}}$ z.B. mit Hilfe des *Gauß-Jordan*-Algorithmus soweit umzuformen,
  - daß die obere linke Teilmatrix $E\left\{\vec{x}\vec{x}^T\right\}$ zu einer Identitätsmatrix wird
  - die untere linke Teilmatrix $E\left\{\vec{y}\vec{x}^T\right\}$ zu einer Nullmatrix wird
  - die obere rechte Teilmatrix ist dann $\underline{A}$
- bei dieser Umformung kann man auch elegant das Problem der linearen Abhängigkeiten der Merkmalskomponenten umgehen (siehe nächsten Abschnitt).

7.4 Merkmalsauswahl

- gibt es lineare Abhängigkeiten, kann $E\left\{\vec{x}\vec{x}^T\right\}$ nicht invertiert werden
  $\rightarrow$ Auswahl einer Teilmenge von linear unabhängigen Merkmalen
- hierzu betrachten wir folgende (Zwischen)Situation



$$\vec{x} = \begin{pmatrix} \tilde{\vec{x}} \\ \vec{u} \end{pmatrix}$$

wobei die Merkmale in $\tilde{\vec{x}}$ bereits ausgewählt wurden

- wir tun so als wollten wir aus $\tilde{\vec{x}}$ jetzt $\vec{u}\vec{y}$ schätzen
  für die Kompmenten aus $\tilde{\vec{x}}$ sei die Matrix $\underline{M}_{\vec{z}}$ schon richtig umgeformt worden:

$$\underline{M}_{\vec{z}} \rightsquigarrow \begin{pmatrix} \mathbb{I} & \underline{\tilde{A}} \\ \mathbb{O} & E\left\{\Delta\tilde{\vec{d}}\Delta\tilde{\vec{d}}^T\right\} \end{pmatrix}$$

- auf der Digonalen von $E\left\{\Delta\tilde{\vec{d}}\Delta\tilde{\vec{d}}^T\right\}$ stehen genau die Var$\{\Delta u_i\}$,
  wobei $\Delta u_i = \tilde{\vec{a}}_i^T \tilde{\vec{x}} - u_i$, also der mittlere Fehler beim Schätzen von u_i aus $\tilde{\vec{x}}$
  (und ebenso – weiter rechts unten – die Var$\{\Delta y_i\}$)
- es gilt: Var$\{\Delta u_i\} = 0 \Leftrightarrow u_i$ und $\tilde{\vec{x}}$ sind linear abhängig
  (d.h. die Schätzung ist perfekt)
  $\Rightarrow$ eliminiere u_i, falls Var$\{\Delta u_i\} < \epsilon$

Pivot-Strategie bei der Reihenfolge des Akzeptierens, um möglichst “gute” Merkmale zu erhalten
(d.h. kleine Anzahl von Merkmalen und kleines S^2)
anstatt nur u_i wie oben zu eliminieren

1. **maximale lineare Unabhängigkeit**
   - wähle u_i mit $i = \operatorname*{argmax}_j \text{Var}\{\Delta u_j\}$,
     d.h. dasjenige u_i, das von den bereits ausgewählten $\tilde{\vec{x}}$ maximale unabhängig
   - wenn $\text{Var}\{\Delta u_i\} < \epsilon$ dann Abbruch
   - gewährleistet gutes numerisches Verhalten, da stets das größte Diagonalelement verwendet wird

2. **maximale Varianz des Fehlers**

- wähle u_i, das am meisten zum Schätzen von $\vec{y}$ beiträgt (was wir ja eigentliche wollen)
  o.B.d.A: wähle u_1 (sonst Umsortieren der Matrix)

$$\left(\begin{array}{cccc} \mathbb{I} & & \underline{\tilde{A}} & \\ & u_1 \;\ldots & \boxed{\vec{b}_1^T} & \\ & \vdots & \ldots & \\ \mathbb{O} & \boxed{\vec{b}_1} \;\ldots & \boxed{E\left\{\Delta\vec{d}\Delta\vec{d}^T\right\}} & \end{array}\right) \rightarrow \left(\begin{array}{cccc} \mathbb{I} & & \underline{\tilde{A}} - \tilde{\vec{a}}_1 \frac{\vec{b}_1^T}{u_1} & \\ & 1 \;\ldots & \boxed{\frac{\vec{b}_1^T}{u_1}} & \\ & 0 & \ldots & \\ \mathbb{O} & \vdots & & \\ & 0 \;\ldots & \boxed{E\left\{\Delta\vec{d}\Delta\vec{d}^T\right\} - \vec{b}_1 \frac{\vec{b}_1^T}{u_1}} & \end{array}\right)$$

  - Zeile mit u_1 wird durch u_1 dividiert
  - auf die j-te Zeile in $\vec{b}_1$ wird das $\frac{b_{1j}}{u_1}$-fache der Zeile, die u_1 enthält, subtrahiert (analog die anderen Zeilen)

das liefert:

- $\Delta S^2 =$ spur $\left(E\left\{\Delta\vec{d}\Delta\vec{d}^T\right\}\right) -$ spur $\left(E\left\{\Delta\vec{d}\Delta\vec{d}^T\right\} - \frac{\vec{b}_1\vec{b}_1^T}{u_1}\right) =$ spur $\left(\frac{\vec{b}_1\vec{b}_1^T}{u_1}\right) = \frac{\vec{b}_1^T\vec{b}_1}{u_1}$
- d.h. wähle u_i mit

  (a) $i =$ argmax$\frac{\vec{b}_i^T\vec{b}_i}{u_i}$ und

  (b) Var$\{\Delta u_i\} < \epsilon$

  vertausche für u_1 und u_i die Zeilen und Spalten
- attraktiv, da direkt mit Schätzfehler verknüpft
- ist natürlich suboptimal, da greedy (wie 1. auch)

Zusammenfassung

- bei linearen Abhängigkeiten der Elemente aus $\vec{x}(\vec{c})$ erhalten wir eine (von mehreren) Lösungen
- die gewählten Elemente von $\vec{x}(\vec{c})$ geben uns ihre “Wichtigkeit”

/* **Berechnung der Matrix $\underline{A}$ für den Polynomklassifikator** */

berechne aus klassif. Stichprobe die Matrix $\underline{M} = \begin{pmatrix} E\{\vec{x}\vec{x}^T\} & E\{\vec{x}\vec{y}^T\} \\ E\{\vec{y}\vec{x}^T\} & E\{\vec{y}\vec{y}^T\} \end{pmatrix}$

FOR alle Zeilen $i = 1, \ldots,$ Dim$(\vec{x}(\vec{c}))$

- VListe[i] := i

FOR alle Zeilen $i = 1, \ldots,$ Dim$(\vec{x}(\vec{c}))$

- bestimme unter den Zeilen $i, \ldots,$ Dim$(\vec{x}(\vec{c}))$ diejenige Zeile k mit $Diag[k] > \varepsilon$ und zugleich $\Delta S^2 = (\vec{b}_k^T \vec{b}_k)/Diag[k]$ ist maximal
- vertausche jeweils Zeilen und Spalten $i \leftrightarrow k$
- zwi := VListe[i]; VListe[i] := VListe[k]; VListe[k] := zwi
- dividiere Zeile i durch $Diag[i]$, d.h. Diagonalelement wird zu 1
- FOR alle Zeilen $k = 1, \ldots,$ Dim$(\underline{M}) \wedge k \neq i$
  - normiere alle Elemente der i-ten Spalte zu 0 (außer $Diag[i]$), d.h.
    k-te Zeile := k-te Zeile - M[k][i] $*$ normierte i-te Zeile
- IF $\Delta S^2 <$ `AbbruchSchranke` **oder** keine Zeile wählbar
  THEN STOP: aktuelle Matrix $\underline{A}$ und VListe ausgeben

7.5 Eigenschaften der Lösung

- Schätzung von $\vec{y}(\vec{c})$ ist "unbiased", d.h.

$$E\left\{\vec{y}(\vec{c}) - \vec{d}(\vec{c})\right\} = 0$$

- $\vec{d}(\vec{c})$ summiert auf 1:

$$\sum d_\kappa(\vec{c}) = 1$$

zur Erinnerung:	für uneingeschränkte Form von $\vec{d}(\vec{c})$ erhalten wir eine Schätzung der a posteriori Wahrscheinlichkeit $p(\omega_\kappa \mid \vec{c})$, die natürlich auf 1 summiert)
wir können	also die $d_\kappa(\vec{c})$ nicht nur zum Klassifizieren verwenden, sondern erhalten auch Konfidenz für die Entscheidung
übrigens	im Allgemeinen gilt **nicht**

$$0 \leq d_\kappa(\vec{c}) \leq 1$$

- $E\left\{(\vec{y}(\vec{c}) - \vec{d}(\vec{c}))^2\right\}$ versus $E\left\{(\vec{p}(\vec{c}) - \vec{d}(\vec{c}))^2\right\}$ versus

  mit $p_\kappa(\vec{c}) = p(\omega_\kappa \mid \vec{c})$

  liefert **identische** Lösung $\underline{A}$

  $p(\omega_\kappa \mid \vec{c})$ wird als “soft labelling” bezeichnet

  - für sehr große Stichproben identischer Klassifikator
  - für kleine Stichproben u.U. nützlich,
    aber “soft labelling” ist aufwendiger als normales Hand-Klassifizieren

7.6 Konfidenzabbildung

- wir erinnern uns, daß wir beim Bayes-Klassifikator als Modifikation mehrere Klassen mit
$$\sum_\kappa p(\omega_\kappa \mid \vec{c}) \geq \theta$$
als Ergebnis liefern können
- auch beim Polynomklassifikator gilt: $\sum d_\kappa(\vec{c}) = 1$, aber $d_\kappa(\vec{c}) \in \mathbb{R}$
- Clipping auf $d_\kappa(\vec{c}) \in [0; 1]$,
ist nicht gut, es ist “unklar”, was wir hierbei machen
- daher

$$\vec{c} \rightarrow \boxed{\text{Klassifikator}} \rightarrow \vec{d}(\vec{c}) \rightarrow \boxed{\text{Konfidenzabbildung}} \rightarrow \tilde{\vec{d}}(\vec{c})$$

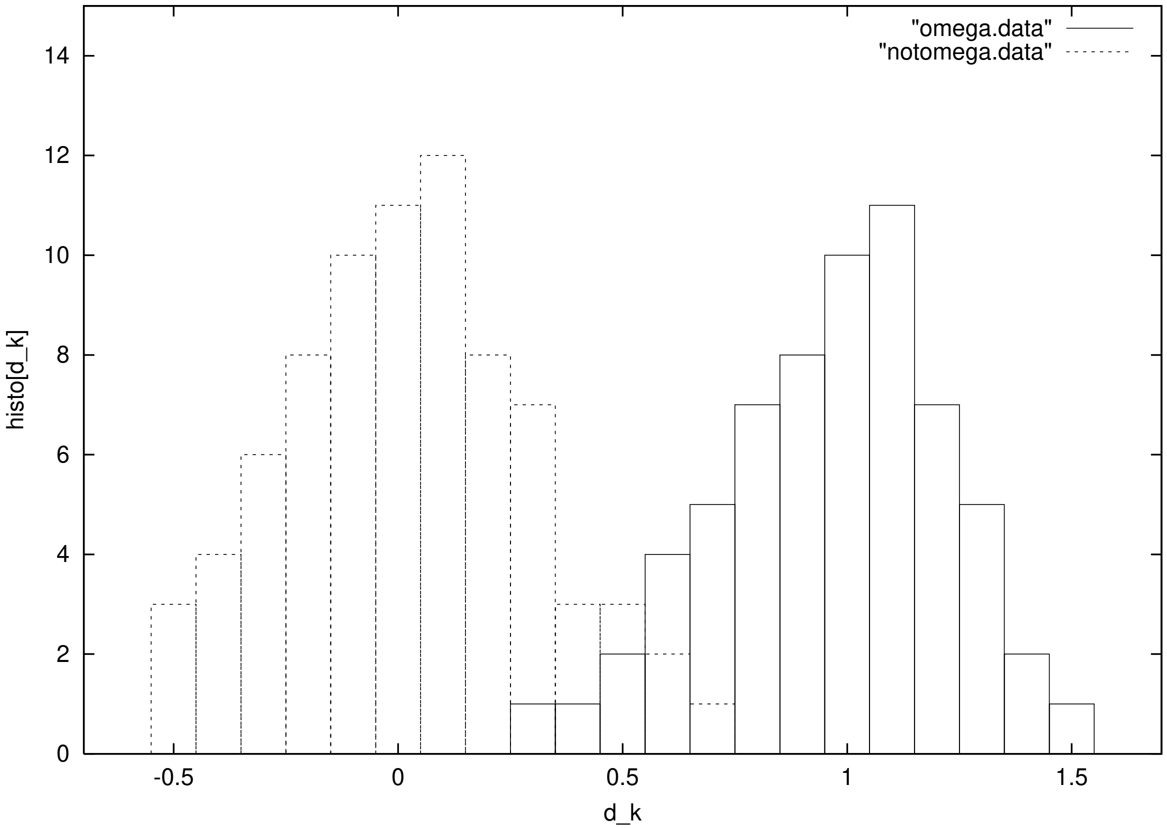
betrachte jedes $d_\kappa(\vec{c})$ unabhängig:

  – $d_\kappa(\vec{c})$ approximiert $y_\kappa(\vec{c}) = \begin{cases} 1 & \vec{c} \in \omega_\kappa \\ 0 & \vec{c} \notin \omega_\kappa \end{cases}$

– wir können daher $d_\kappa(\vec{c})$ als eindimensionalen Merkmal(svektor) auffassen für das zwei Klassenproblem:

$$\vec{c} \in \omega_\kappa \text{ vs. } \vec{c} \notin \omega_\kappa$$

Beispiel:



- da wir jetzt im eindimensionalen Fall sind, können wir
  - mit relativen Histogramme approximieren wir:

$$p(d_\kappa(\vec{c}) \mid \omega_\kappa) \text{ und } p(d_\kappa(\vec{c}) \mid \neg\omega_\kappa)$$

  - durch relative Häufigkeiten schätzen wir:

$$p(\omega_\kappa) \text{ und } p(\neg\omega_\kappa)$$

  - damit Konfidenz für ω_κ aus den absoluten Histogrammen histo:

$$\begin{aligned}
\text{cnf}(d_\kappa) &= \frac{p(\omega_\kappa, d_\kappa(\vec{c}))}{p(\omega_\kappa, d_\kappa(\vec{c})) + p(\neg\omega_\kappa, d_\kappa(\vec{c}))} \\
&= \frac{p(d_\kappa \mid \omega_\kappa)p(\omega_\kappa)}{p(d_\kappa \mid \omega_\kappa)p(\omega_\kappa) + p(d_\kappa \mid \neg\omega_\kappa)p(\neg\omega_\kappa)} \\
&\sim \frac{\text{histo}_{\omega_\kappa}[d_\kappa] \cdot \hat{P}(\omega_\kappa)}{\text{histo}_{\omega_\kappa}[d_\kappa] \cdot \hat{P}(\omega_\kappa) + \text{histo}_{\neg\omega_\kappa}[d_\kappa] \cdot \hat{P}(\neg\omega_\kappa)}
\end{aligned}$$

- **aber** Summation der cnf auf 1 geht verloren

  unter der Annahme, daß die Merkmale tatsächlich aus (genau) einem stochastischen Prozeß der K Klassen stammt (*closed world assumption*)

  explizite Normalisierung:

$$\tilde{\vec{d}}(\vec{c}) = \frac{1}{\sum_{\kappa=1}^{K} \mathsf{cnf}(d_\kappa(\vec{c}))} \begin{pmatrix} \mathsf{cnf}(d_1(\vec{c})) \\ \vdots \\ \mathsf{cnf}(d_K(\vec{c})) \end{pmatrix}$$

  (hierbei müssen die $d_\kappa(\vec{c})$ durch die entsprechende Diskretisierung der Histogrammbildung ersetzt werden)

- falls ein Muster aus **keiner** der K Klassen entstammt (d.h. Widerspruch zur *closed world assumption*), kann dies an den Histogrammen/Dichten

$$p(d_\kappa \mid \omega_\kappa), p(d_\kappa \mid \neg\omega_\kappa)$$

  für **alle** κ erkannt werden