2.2.2) Diskriminanzanalyse

Schon gesehen: die Logistische Regression eignet sich besonders bei Zielgrößen, die binär sind

$$Y \begin{cases} \rightarrow \text{"ja"} & (1) \\ \rightarrow \text{"nein"} & (0) \end{cases}$$

Im Fall von Y mit mehreren Ausprägungen sind andere Methoden besser geeignet → zB. Diskriminanzanalyse (klass. Verfahren)

Annahmen:

- jedes Objekt gehört zu genau einer von mehreren potentiellen Klassen
- es gibt bereits zu jeder Klasse Objekte, deren Klassenzugehörigkeit bekannt ist (supervised)

1 Einführendes Beispiel & Grundidee

Bsp.: Für einen hochwertigen Laptop soll personalisierte Werbung durchgeführt werden. Am besten an Kunden, die auch prinzipiell Interesse haben

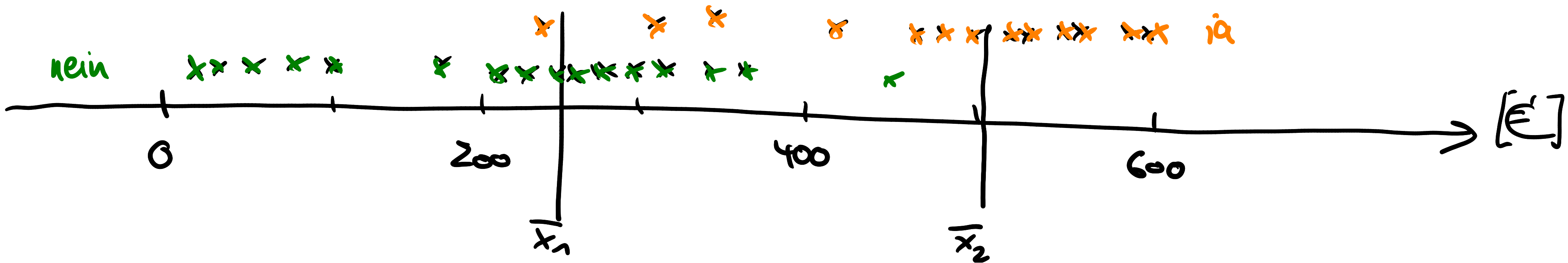
FRAGE: "Hätten Sie prinzipiell Interesse?"
"Wieviel haben Sie im letzten Jahr für Elektronik ausgegeben?"

		Betrag [€]					
x^1	kein Interesse	190	480	230	110	320	...
x^2	Interesse	850	420	420	480	360	...

20 Werte, $n_1 = 20$; $\overline{x^1} = 246$ (MW)

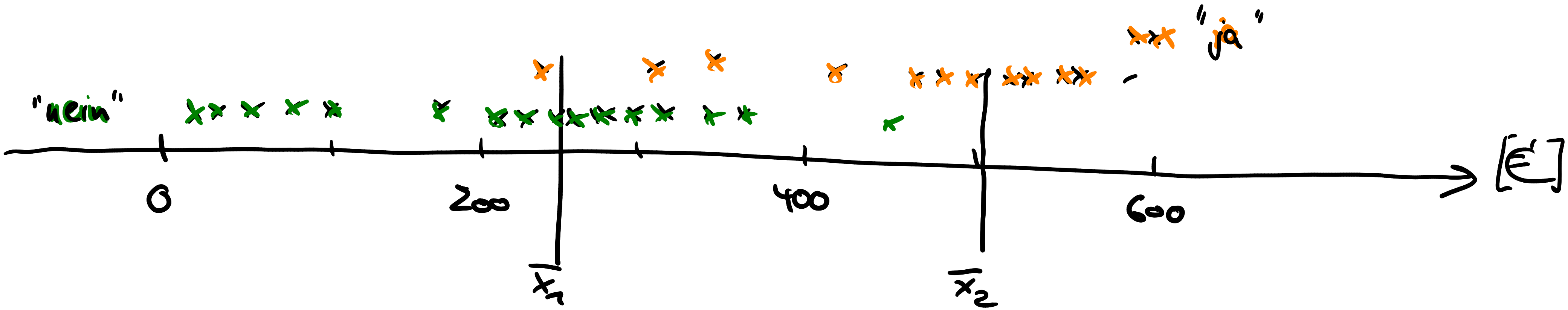
20 Werte, $n_2 = 20$; $\overline{x^2} = 504$ (MW)

einfachste Idee: benutze die Mittelwerte der Gruppen um eine Entscheidung zu treffen!



x: kein Interesse
x: Interesse

⇒ Idee: Kunden, die kein Interesse am Gerät haben geben im Mittel weniger für Elektronik aus.



⇨ Bei gleicher (ähnlicher) Streuung des Merkmals "Ausgaben in €" in beiden Gruppen könnte man so entscheiden:

Falls $\boxed{|x-\bar{x}^1| < |x-\bar{x}^2| \Rightarrow x \text{ in Gruppe } 1}$ sonst 2

⇒ kleinster Abstand zum Gruppenmittel entscheidet!
Geht immer ☺

ABER: ziemlich fehleranfällig, wenn die Gruppen große Überschneidungsbereiche haben.

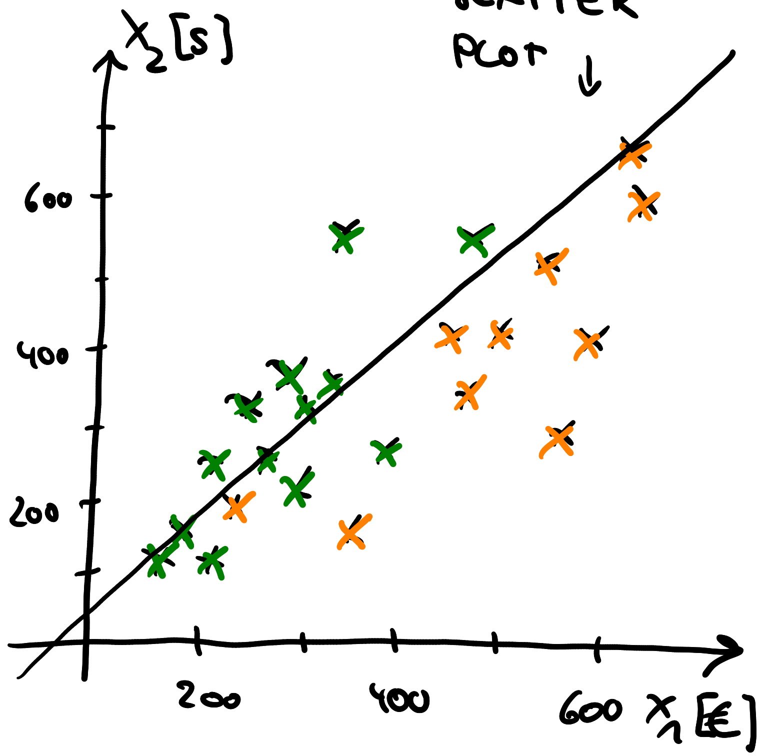
⇒ weiteres zusätzliches Merkmal

"Wieviel Zeit (s) verbringen Sie pro Tag auf Websites mit Elektronikprodukten?"

x_1^1	kein Int	190	430	230	110	320	...	← [€]
x_2^1	— " —	160	270	150	60	100	...	← [s]
x_1^2	Inter.	350	420	420	480	360	...	← [€]
x_2^2	— " —	220	190	70	210	150		← [s]

⇒

SCATTER PLOT



kein Interesse

Interesse

IDEE ⇒ Finde eine "optimale" Gerade, die die Objekte trennt!

→ aber: was ist optimal??

... vielleicht kann man die erste Idee mit den Mittelwerten übertragen?

IDEE Eine Linearkombination der beiden Merkmale ergibt eine neue Variable

<u>Ansatz</u>:

$$\boxed{y_j = a_1 x_{j1} + a_2 x_{j2}}$$

$j = 1, \dots, n$ Datenpunkte

$\leftarrow$ Koeffizienten a_1, a_2

$\Rightarrow$ Wähle die Koeffizienten a_1, a_2 so, daß die neue Größe y eine möglichst große Streuung zwischen beiden Gruppen und eine möglichst kleine Streuung innerhalb der Gruppen besitzt! $\Rightarrow$ separiert die Gruppen optimal

Mit optimalen a_1, a_2 gilt:

$$\boxed{|y - \overline{y^1}| < |y - \overline{y^2}| \Rightarrow (x_1, x_2) \text{ in Gruppe } ① \quad (*)}$$

② sonst

dabei ist
- $x = (x_1, x_2)$ neues Objekt das zugeordnet werden soll
- $y = a_1 x_1 + a_2 x_2$ neues Objekt in neuer Variablendarstellung
- $\overline{y}_1$ Mittelwert der $y_j = a_1 x_{j1} + a_2 x_{j2}$ in Gruppe 1
- $\overline{y}_2$ Mittelwert der $y_i = a_1 x_{j1} + a_2 x_{j2}$ in Gruppe 2

② Bestimmung der Koeffizienten a_1, a_2

Betrachte s_w^2 : Streuung der y-Werte innerhalb der Gruppen

s_b^2 : Streuung der y-Werte zwischen den Gruppen

"optimal" $\hat{=}$ s_b^2 groß, s_w^2 klein $\Rightarrow$ maximiere $\boxed{\lambda = \frac{s_b^2}{s_w^2}}$ ← Quotient der Streuungen wir brauchen also s_b^2 und s_w^2 um eine Darstellung für λ zu finden ...

EXKURS: Streuungszerlegung

Allgemein: Stichprobe mit Umfang n, k Gruppen

- k Gruppen mit $n_1, n_2, \dots, n_k$ Elementen
  $\overline{x}_1, \overline{x}_2, \dots, \overline{x}_k$ Mittelwerte der Gruppen
  $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$ Varianzen der Gruppen

dann gilt:

$$\boxed{\begin{aligned} n &= n_1 + n_2 + \dots + n_k \\ \overline{x} &= \frac{1}{n}\left(n_1\overline{x}_1 + n_2\overline{x}_2 + \dots + n_k\overline{x}_k\right) \end{aligned}}$$

dann ist die externe Varianz die Varianz der Gruppenmittelwerte

EXTERNE VARIANZ →

$$\boxed{\sigma_{EXT}^2 = \frac{1}{n}\left(n_1(\bar{x}_1-\bar{x})^2+\dots+n_k(\bar{x}_k-\bar{x})^2\right)}$$

und die interne Varianz beschreibt die Variabilität innerhalb der Gruppen

$$\boxed{\sigma_{INT}^2 = \frac{1}{n}\left(n_1\sigma_1^2+\dots+n_k\sigma_k^2\right)}$$

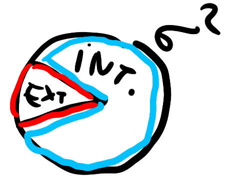
und es gilt:

$$\boxed{\sigma^2 = \sigma_{int}^2 + \sigma_{ext}^2}$$

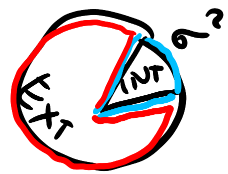
← STREUUNGSZERLEGUNG

INTERPRETATION → HOMOGENITÄT

- kleine Untersch. zwischen den Gruppen → σ_{int}^2 groß, σ_{ext}^2 klein



- große Untersch. zwischen den Gruppen → σ_{int}^2 klein, σ_{ext}^2 groß



Bei uns ist die Variable für die wir uns interessieren

$$\boxed{y_j = a_1 x_{j1} + a_2 x_{j2}}$$

$j = 1, \dots, n$

Also gilt:

$$\boxed{s_w^2 = (a_1, a_2)\, C_w \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}} \qquad \boxed{s_b^2 = (a_1, a_2)\, C_b \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}}$$

wobei C_b: Kovarianzmatrix von X_1, X_2 zwischen den Gruppen

C_w: Kovarianzmatrix von X_1, X_2 innerhalb der Gruppen

Wie sehen in unserem Beispiel C_b und C_w konkret aus?

KLAR: Für 2 Features X_1, X_2 und 2 Gruppen müssen beide Matrizen 2×2 sein...

Sei

$\overline{x_1}, \overline{x_2}$	Mittelwert von X_1, X_2 für alle Daten	$Var^1(X_1), Var^1(X_2)$ $Var^2(X_1), Var^2(X_2)$	Varianzen von X_1, X_2 bzgl. Gruppen 1 und 2
$\overline{x_1^1}, \overline{x_2^1}$	Mittelwert von X_1, X_2 in Gruppe 1 (2 analog)	$Cov^1(X_1,X_2)$ $Cov^2(X_1,X_2)$	Kovarianz von X_1, X_2 bzgl. Gruppen 1 und 2

$$C_b = \begin{bmatrix} \frac{1}{n_1+n_2}\left(n_1\left(\overline{x_1^1}-\overline{x_1}\right)^2 + n_2\left(\overline{x_1^2}-\overline{x_1}\right)^2\right) & \frac{1}{n_1+n_2}\left(n_1\left(\overline{x_1^1}-\overline{x_1}\right)\left(\overline{x_2^1}-\overline{x_2}\right) + n_2\left(\overline{x_1^2}-\overline{x_1}\right)\left(\overline{x_2^2}-\overline{x_2}\right)\right) \\ \frac{1}{n_1+n_2}\left(n_1\left(\overline{x_1^1}-\overline{x_1}\right)\left(\overline{x_2^1}-\overline{x_2}\right) + n_2\left(\overline{x_1^2}-\overline{x_1}\right)\left(\overline{x_2^2}-\overline{x_2}\right)\right) & \frac{1}{n_1+n_2}\left(n_1\left(\overline{x_2^1}-\overline{x_2}\right)^2 + n_2\left(\overline{x_2^2}-\overline{x_2}\right)^2\right) \end{bmatrix}$$

EXTERNE Kovarianzmatrix

(zwischen den Gruppen)

$$C_W = \begin{bmatrix} \frac{1}{n_1+n_2}\left(n_1 Var^1(x_1) + n_2 Var^2(x_1)\right) & \frac{1}{n_1+n_2}\left(n_1 Cov^1(x_1,x_2) + n_2 Cov^2(x_1,x_2)\right) \\ \frac{1}{n_1+n_2}\left(n_1 Cov^1(x_1,x_2) + n_2 Cov^2(x_1,x_2)\right) & \frac{1}{n_1+n_2}\left(n_1 Var^1(x_2) + n_2 Var^2(x_2)\right) \end{bmatrix}$$

INTERNE

Kovarianzmatrix

(innerhalb der Gruppen)

Mit

$$\boxed{S_W^2 = (a_1, a_2)\, C_W \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}} \qquad \boxed{S_b^2 = (a_1, a_2)\, C_b \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}}$$

gilt also:

$$\boxed{\lambda = \frac{S_b^2}{S_W^2} = \frac{(a_1, a_2)\, C_b \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}}{(a_1, a_2)\, C_W \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}} \longrightarrow \max_{a_1, a_2} (\ldots)}$$

Maximierungs-problem!

λ_{max} gesucht, so daß

$$\boxed{C_b \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} - \lambda C_W \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = 0} \Leftrightarrow \boxed{(C_b - \lambda C_W)\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = 0} \quad (*)$$

falls C_w vollen Rang hat $\Rightarrow$ C_w ist invertierbar (X_1, X_2 unabhängig, bzw nicht kollinear!)

(*) $\Rightarrow$ $$\boxed{\left(C_w^{-1} C_b - \lambda\, Id\right)\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = 0}$$ $\leftarrow$ Eigenwertproblem

$\Rightarrow$ Suche Eigenvektor $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ zum größten Eigenwert λ!

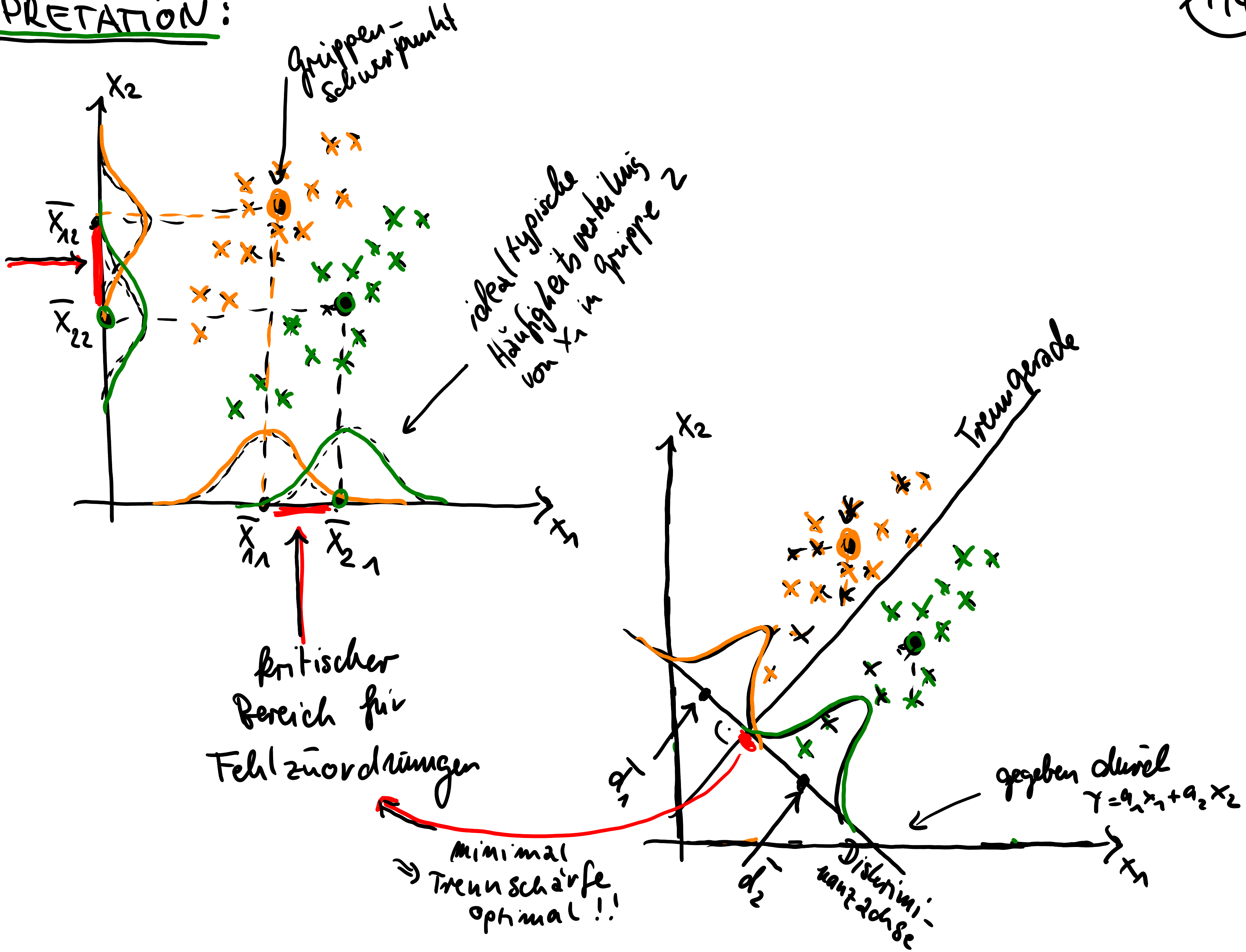
dann $\boxed{y = a_1 x_1 + a_2 x_2}$ neue Variable mit Entscheidungskriterium

$$\boxed{|y - \bar{y}^1| < |y - \bar{y}^2| \Rightarrow (x_1, x_2) \text{ in Gruppe } ①}$$

Gruppe ② sonst

Die Lösung des Eigenwertproblems liefert uns also genau die neue Variable y, für die das Entscheidungskriterium die beste Trennschärfe für die beiden Gruppen besitzt

INTERPRETATION:



Allgemeine Formulierung

LINEARE DISKRIMINANZANALYSE (LDA)

AUSGANGSLAGE

→ Gesamtheit mit r Gruppen (Zuordnung bekannt)

→ k unterschiedliche Merkmale (= Features)

→ jeweils n_i $(i=1,\dots,r)$ Beobachtungen im Gruppe i

→ beliebige (neue) Beobachtung $x=(x_1,\dots,x_k)$

→ j-te Beobachtung in Gruppe i $\quad x_j^i=(x_{j_1}^i,\dots,x_{j_k}^i)$ $\quad (j=1,\dots,n_i)$

1. lineare Diskriminanzanalyse ⇒ Verallgemeinerung des vorigen Beispiels

2. Erweiterung: Zuordnung mit Hilfe kanonischer Variablen

I. Aufgabenstellung

Die lineare Diskriminanzanalyse basiert auf der (linearen)

DISKRIMINANZFUNKTION $\boxed{y = a_1 x_1 + \dots + a_k x_k}$

mit Gruppenmittelwerten $\boxed{\bar{y}^i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_j^i}$

Für einen Punkt $x = (x_1, \dots, x_k)$ mit unbekannter Zuordnung gilt dann für

$$\boxed{y = a_1 x_1 + \dots + a_k x_k}$$

$\boxed{x \text{ gehört zu derjenigen Gruppe } i \text{ für die } |y - \bar{y}^i| \text{ minimal ist } (i = 1, \dots, r)}$

dabei gilt: Die Diskriminanzfunktion soll so gewählt werden, daß sie möglichst trennscharf ist, d.h.

für
$$s_w^2 = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_j^i - \overline{y^i})^2$$
$$s_b^2 = \frac{1}{n}\sum_{i=1}^{r} n_i(\overline{y^i} - \overline{y})^2$$
$$\boxed{\frac{s_b^2}{s_w^2} = \lambda}$$ maximal

II. LÖSUNG DES MAXIMIERUNGSPROBLEMS

Sei $\overline{x} = (\overline{x_1}, \ldots, \overline{x_k})$ ← Zeilenvektor der Mittelwerte der k Merkmale (alle Gruppen)

$\overline{x^i} = (\overline{x_1^i}, \ldots, \overline{x_k^i})$ ← Vektoren der Mittelwerte der k Merkmale in den r Gruppen $(i=1,\ldots,r)$

$$C^i = \frac{1}{n_i}\sum_{j=1}^{n_i}(x_j^i - \overline{x^i})^t(x_j^i - \overline{x^i})$$ Kovarianzmatrix innerhalb der i-ten Gruppe $(i=1,\ldots,r)$

Dann gilt für $C_w = \frac{1}{n}\sum_{i=1}^{k} n_i C^i$ und $C_b = \frac{1}{n}\sum_{i=1}^{k} n_i (\bar{x^i} - \bar{x})^t (\bar{x^i} - \bar{x})$

sowie $\vec{a}^t = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ Koeffizientenvektor $\vec{a} = (a_1, \dots, a_n)$

$$J = \frac{\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} C_b (a_1 \dots a_n)}{\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} C_w (a_1, \dots, a_n)} \Leftrightarrow \boxed{C_b \vec{a} = \lambda C_w \vec{a}}$$

$B\vec{a} = \lambda W \vec{a}$

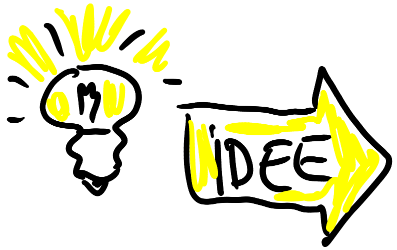
verallgemeinertes symmetrisches Eigenwertproblem

$$\Leftrightarrow \boxed{(C_w^{-1} C_b - \lambda \cdot Id)\vec{a}^t = 0}$$

$\Rightarrow$ Der gesuchte Vektor $\vec{a}$ entspricht dem Eigenvektor zum größten Eigenwert λ!

$\Rightarrow$ EV normiert bzgl. W (C_w): $a^t W a = 1$

III. Güte der Diskriminanzfunktion



Betrachte die erklärte Streuung
= Streuung zwischen den Gruppen

$\boxed{\lambda = \frac{S_b^2}{S_w^2}}$, dann $\frac{\text{erklärte Streuung}}{\text{Gesamtstreuung}} = \frac{S_b^2}{S_w^2 + S_b^2} = \frac{\lambda}{1+\lambda}$

⇒ Gütemaß: **kanonischer Korrelationskoeffizient** $\boxed{c = \sqrt{\frac{\lambda}{1+\lambda}}} \in [0;1]$

"je größer desto besser ☺"

manchmal auch andersrum $\boxed{\Lambda = \frac{1}{1+\lambda}}$ **WILKS-LAMBDA** $\in [0;1]$

"Anteil der nicht-erklärten Streuung an der Gesamtstreuung" ⇒ top ist $\Lambda = 0$!

Bemerkung:

Vorteile der LDA:

- intuitiv verständlich
- analytische Lösung
- keine Meta-Parameter
- kann als Dimensionsreduktion interpretiert werden

$\Rightarrow y = a_1 x_1 + \dots + a_n x_n$ läßt sich visualisieren!

UND: LDA läßt sich so erweitern, daß auch weitere Eigenwerte und -vektoren des EWP $\boxed{C_b \vec{a} = \lambda C_w \vec{a}}$ mit berücksichtigt werden!

VORTEIL: auch die restliche Streuung der Merkmale wird bei der Klassifikation berücksichtigt $\Rightarrow$ mehr Trennschärfe!!

Ausblick: Zuordnung mit Hilfe kanonischer Variablen

Sind $\vec{a}^1, \dots, \vec{a}^q$ Eigenvektoren zu den positiven EW $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_q$

so lassen sich für eine beliebige Beobachtung $x = (x_1, \dots, x_k)$

die **KANONISCHEN VARIABLEN** $\boxed{y_j = \bar{x}\,\vec{a}^i}$ $(i = 1, \dots, q)$ definieren.

Für die kanonischen Variablen der Gruppenmittel $\bar{x}^i$ (Vektoren) schreibt man $\overline{y_j}^i$ $(i = 1, \dots, r;\ j = 1, \dots, q)$

Ein Objekt gehört dann zu der Gruppe, für die $\boxed{\sum_{i=1}^{q} (y_i - \overline{y}_j^i)^2}$ minimal ist.