LDA – mehr als 2 Gruppen

Obwohl die Klassifikation bei mehr als zwei Gruppen auch paarweise vorgenommen werden könnte, gibt uns dieser Ansatz keine geeignete graphische Darstellung der Ergebnisse.

Für 3 Gruppen sollten wir eine 2d Projektion finden, die die Gruppen optimal trennt.

Dies kann über einen Vergleich der Varianz zwischen den Gruppen und der Varianz in den Gruppen (wie in der MANOVA) passieren:

$$\mathbf{B} = \sum_{i=1}^{g} n_i(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})', \quad \mathbf{W} = \sum_{i=1}^{g} (n_i - 1)\mathbf{S}_i$$

wobei g die Zahl der Gruppen, und $\bar{\mathbf{X}}$ der Gesamtmittelwert ist.

... mehr als 2 Gruppen

Als Lösung für $\hat{\mathbf{a}}'$, welches

$$\frac{\hat{\mathbf{a}}'\mathbf{B}\hat{\mathbf{a}}}{\hat{\mathbf{a}}'\mathbf{W}\hat{\mathbf{a}}}$$

maximiert, ergibt sich als der erste Eigenvektor $\mathbf{e}_1$ von $\mathbf{W}^{-1}\mathbf{B}$.

Der *Diskriminanten Raum* wird durch die Eigenvektoren von $\mathbf{W}^{-1}\mathbf{B}$ aufgespannt. Die erste Diskriminante ist der erste Eigenvektor, ... Für g Gruppen erhält man durch die ersten $(g-1)$ Diskriminanten Richtungen, $\mathbf{e}_j, \quad j = 1, \dots, g-1$ die optimalen Projektionen.

Die Entscheidungsregel wird dann folgendermaßen gebildet: Ordne $\mathbf{x}_0$ Gruppe k zu, falls

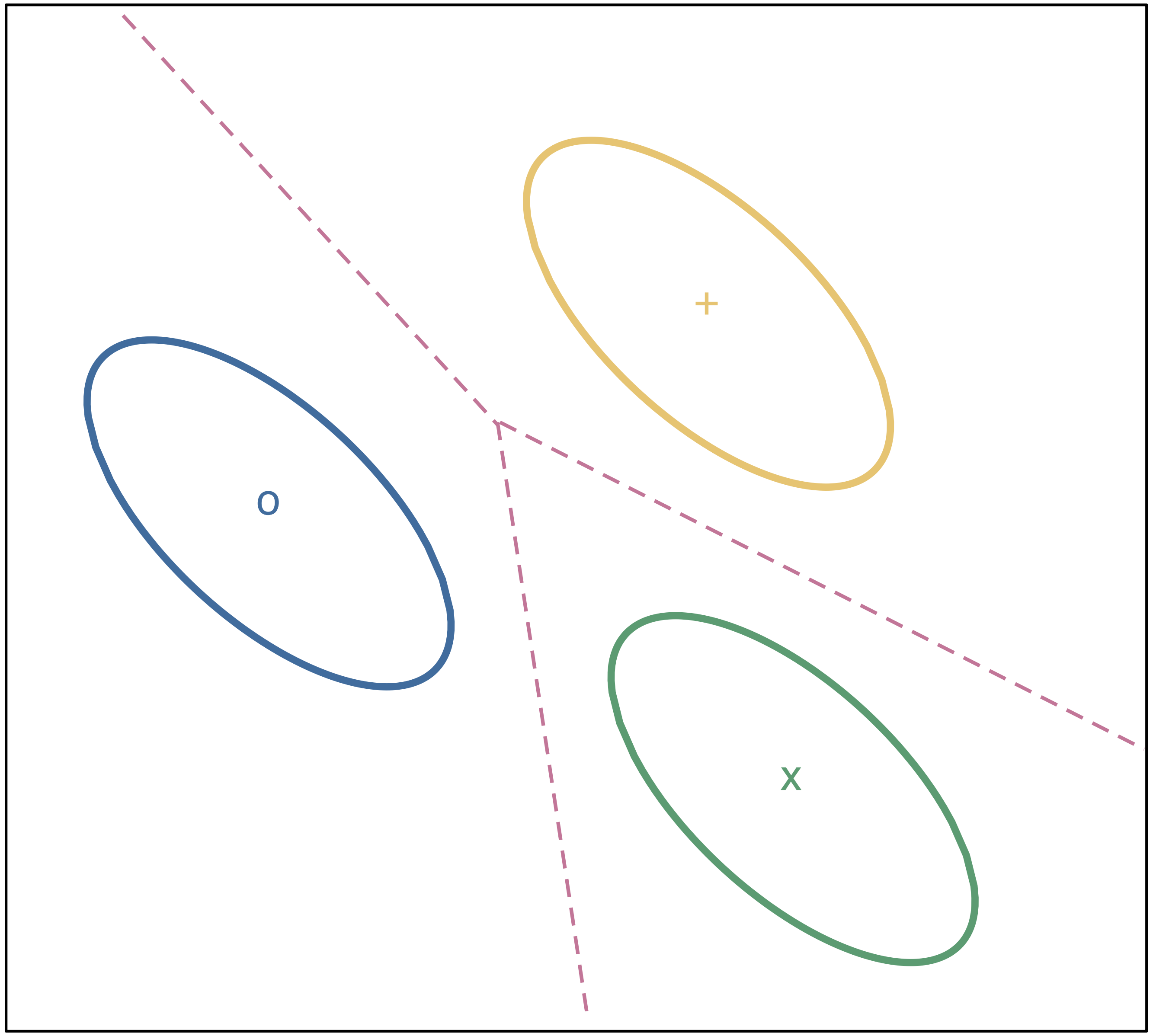
$$\sum_{j=1}^{r} (\hat{\mathbf{e}}_j'(\mathbf{X}_0 - \bar{\mathbf{X}}_k))^2 \leq \sum_{j=1}^{r} (\hat{\mathbf{e}}_j'(\mathbf{X}_0 - \bar{\mathbf{X}}_i))^2, \quad \forall k \neq i$$

wobei $r \leq (g-1)$ die Anzahl der benutzen Diskriminanten ist.

Beispiel: 3 Gruppen

- Am Beispiel von 3 Gruppen in einer 2d-Projektion, lässt sich sehr gut erkennen, dass die Varianz Covarianz Struktur der entscheidende Faktor bei der Zuordnung ist.

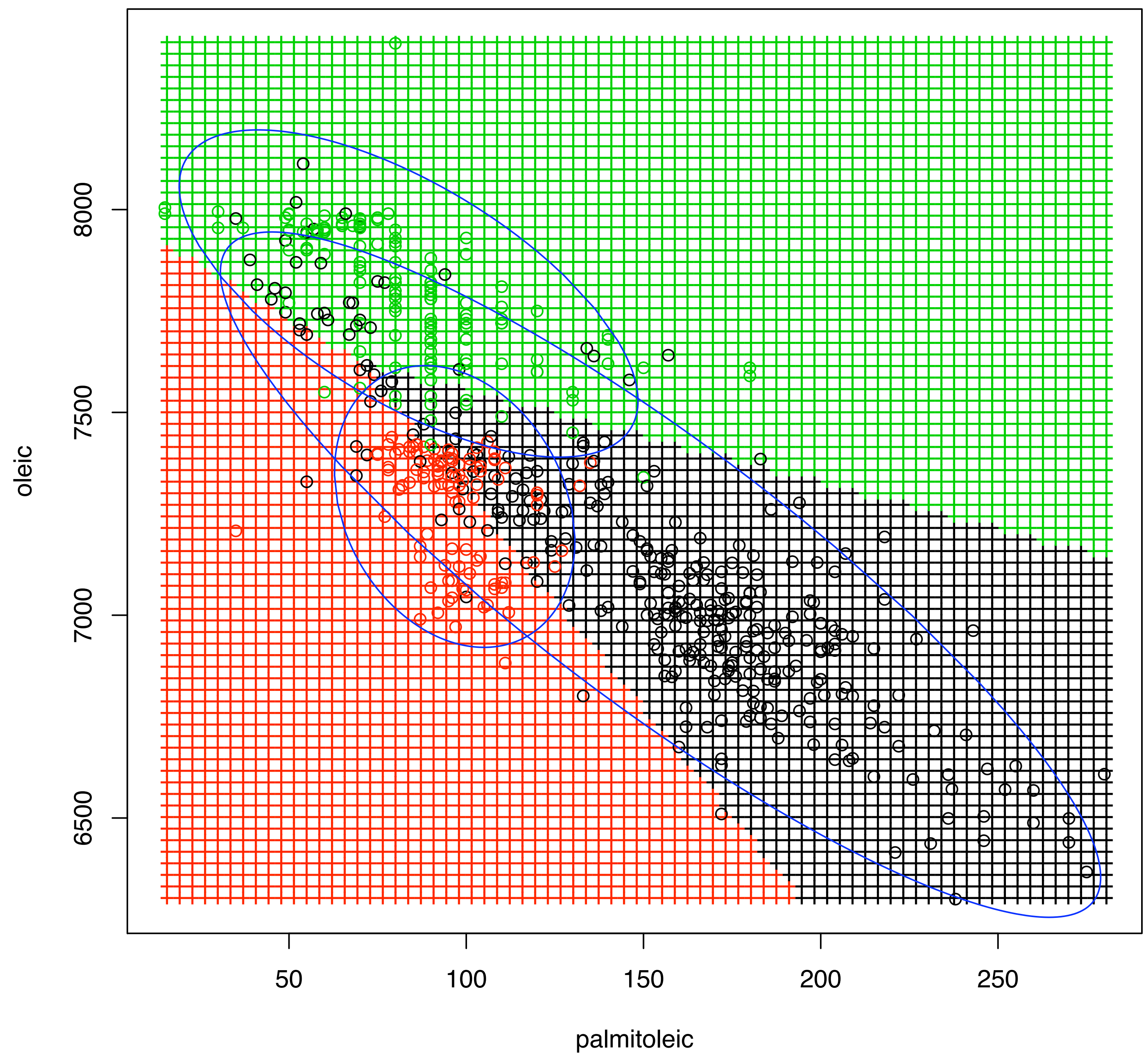
  (Die euklidische Distanz wird durch die Mahalanobis Distanz “ersetzt”)



LDA – Abweichung von gleicher Varianz

- In vielen Fällen ist die Annahme der gleichen Varianz nicht haltbar, so dass die LDA keine sinnvollen Ergebnisse liefern muss.

- Oft ist die Korrelation der Gruppen jedoch ähnlich, so dass sich durch dass “poolen” der Varianzen dennoch akzeptable Resultate ergeben.



Quadratische Diskriminanzanalyse

Quadratische Diskriminanz Analyse bezieht auch die Information der **individuellen** Varianz Covarianz Matrizen ein, im Gegensatz zur LDA darf also gelten $\Sigma_i \neq \Sigma_j$.

Als Regel ergibt sich:
Weise eine Beobachtung $\mathbf{x}_0$ Gruppe 1 zu, falls

$$-\frac{1}{2}\mathbf{x}_0' \left(\mathbf{S}_1^{-1} - \mathbf{S}_2^{-1}\right) \mathbf{x}_0 + \left(\bar{\mathbf{x}}_1' \mathbf{S}_1^{-1} - \bar{\mathbf{x}}_2' \mathbf{S}_2^{-1}\right) \mathbf{x}_0$$

$$-\frac{1}{2}\left(\ln\left(\frac{|\mathbf{S}_1|}{|\mathbf{S}_2|}\right) + \left(\bar{\mathbf{x}}_1' \mathbf{S}_1^{-1} \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2' \mathbf{S}_2^{-1} \bar{\mathbf{x}}_2\right)\right) \quad \geq \quad \ln \frac{c(1|2)}{c(2|1)} \frac{p_2}{p_1}$$

gilt, sonst Gruppe 2.

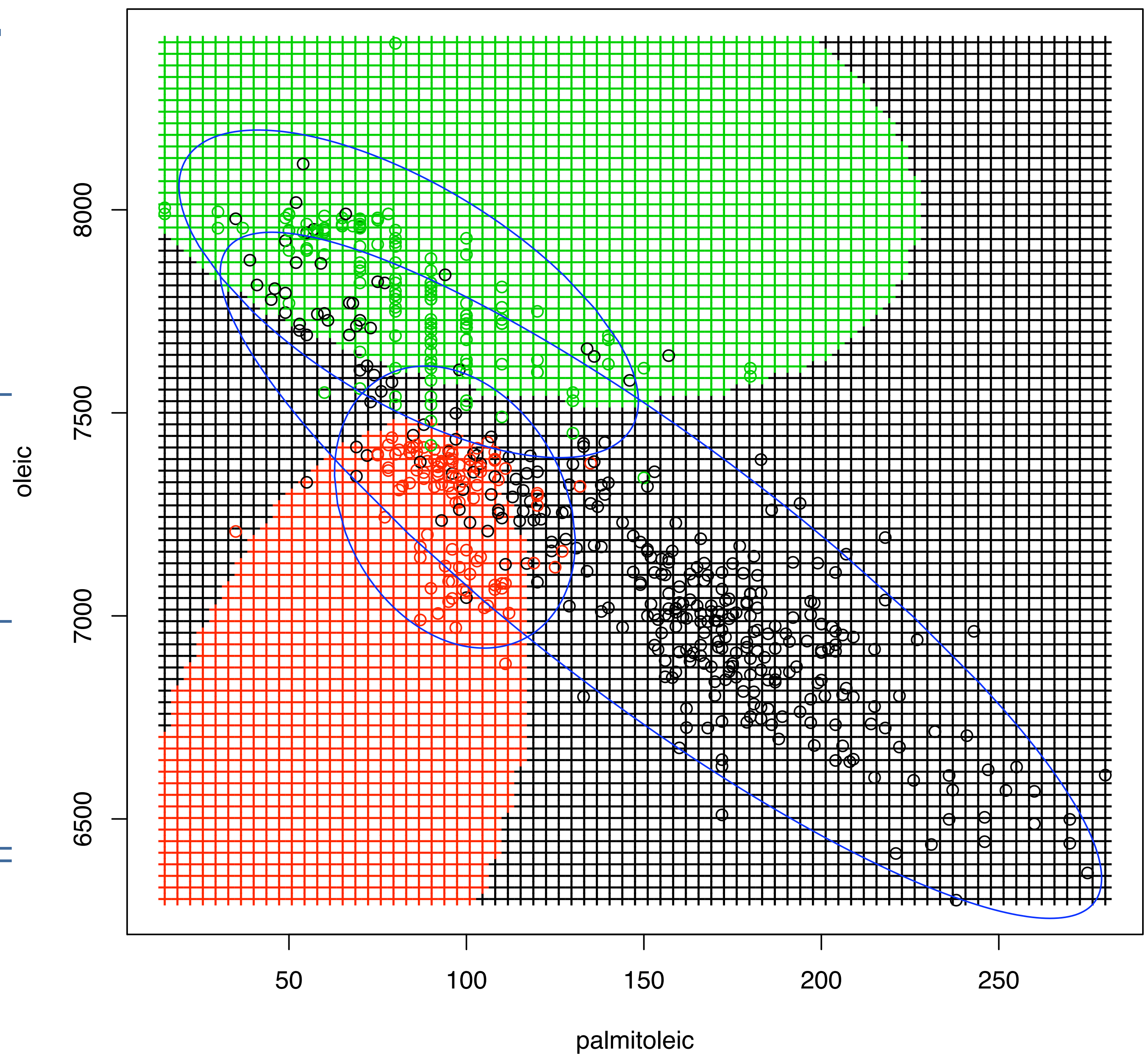
In höheren Dimensionen, und mit mehr als 2 Gruppen entarten die Ergebnisse der QDA leicht.

QDA – Beispiel

- Gleiches Beispiel wie auf Seite 160 – nun mit QDA.

- Konfusionsmatrix:

Vorhers.		Region 1	2	3
LDA	1	270	36	11
	2	18	62	2
	3	35	0	138
QDA	1	262	14	13
	2	27	84	1
	3	34	0	137
Summe		323	98	151



Bemerkungen

- Grundsätzlich lässt sich die Diskriminanz Analyse über zwei Ansätze motivieren:
  - Probabilistisch
    Hier wird versucht, unter Annahme der Normalverteilung, die Beobachtungen entsprechend ihrer zu erwartenden Dichte zuzuordnen.
  - Fishers Ansatz
    Es wird die Linearkombination der Daten gesucht, die die beste Trennung, unter Berücksichtigung der Varianz Covarianz Struktur, zwischen den Gruppen liefert.
- Wenn beide Ansätze die Varianz Covarianz Struktur berücksichtigen sind ihre Ergebnisse äquivalent.
- Oft werden als Input für die LDA auch quadrierte Variablen verwendet. Die Ergebnisse sind mit denen der QDA vergleichbar.

Bewertung der Klassifikationsergebnisse

- Die Bewertung eines Klassifizierers ist verzerrt, wenn sie mit den gleichen Daten vorgenommen wird, die zur Erstellung des Klassifizierers benutzt wurden.
- **Lösung**: Benutze Test- und Trainings-Stichprobe