UE Statistische Mustererkennung
WS 2021
Angaben zur 3ten Aufgabengruppe

1 Aufgabe UE-III.1

1.1

a) Generieren Sie eine Stichprobe vom Umfang 300 für die bivariate Normalverteilung mit Mittelwert $(4,7)^T$ und Varianzen bzw. Korrelationskoeffizienten $\sigma_{11} = 12, \sigma_{22} = 2, \rho_{12} = -0.5$. Gehen Sie dabei wie folgt vor:

   – Berechnen Sie die Eigenwertzerlegung der Kovarianzmatrix, und aus dieser die inverse *whitening*-Transformation.

   – Generieren Sie 600 unabhängig $N(0,1)$ verteilte Zufallswerte, und ordnen Sie diese in einer 2×300-Matrix an; dies entspricht einer Stichprobe vom Umfang 300 einer bivariaten Normalverteilung mit dekorrelierten und z-standardisierten Variablen X, Y.

   – Wenden Sie die inverse *whitening* Transformation auf die obige Stichprobe an.

b) Schätzen Sie aus der Stichprobe das Mittel, die Kovarianzmatrix sowie die Korrelationsmatrix (berechnet sich aus der Kovarianzmatrix, indem man jedes Element durch das Produkt der korrespondierenden Standardabweichungen dividiert).

b) Vergleichen Sie Ihre Ergebnisse mit den wahren Verteilungsparametern aus dem obigen Teilbeispiel, sowie mit der Kovarianz- resp. Korrelationsschätzung des von Ihnen verwendeten Softwarepaktes.

2 Aufgabe UE-III.2

Plotten Sie den *Rayleigh-Quotienten*

$$r(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{A} \mathbf{w}}{\mathbf{w}^T \mathbf{w}} \tag{1}$$

sowohl für die Kovarianz- als auch die Korrelations-Matrix in Beispiel III.1, indem Sie wie folgt vorgehen: lassen Sie den Richtungsvektor $\mathbf{w}$ von 0 bis 360 Grad rotieren und geben Sie die mit dem Wert des Rayleigh-Quotienten skalierten Richtungsvektoren $r(\mathbf{w})\frac{\mathbf{w}}{\|\mathbf{w}\|}$ aus (dies ist eine parametrische Kurve im $\mathbb{R}^2$).

Ergänzen Sie die Darstellung der Kurven durch die mit ihren korrespondierenden Eigenwerten skalierten Eigenvektoren der Kovarianz- bzw. Korrelationsmatrix.

3 Aufgabe UE-III.3

Die Dateien namens *ldaTrain.txt* und *ldaTest.txt* enthalten je 500 2-dimensionale Merkmalsvektoren. Die ersten 200 Punkte gehören zur Klasse ω_1, die nächsten 200 Punkte zur Klasse ω_2 und die letzten 100 zur Klasse ω_3.

a) Lesen Sie die Datei *ldaTrain.txt* ein und fassen Sie die 500 Punkte als Zufallsstichprobe einer *mixture distribution* dreier bivariater Normalverteilungen auf. Trainieren Sie einen linearen Klassifikator unter der Annahme, daß die Merkmale für alle drei Klassen normalverteilt mit identischen Kovarianzmatrizen sind, mittels LDA. Schätzen Sie dazu die *a priori*-Wahrscheinlichkeiten, Klassen-Mittel und die gemeinsame Kovarianzmatrix aus den Daten. Um eine gemeinsame (gepoolte) Kovarianzschätzung für alle Klassen zu erhalten, verwenden Sie

$$\hat{\mathbf{C}} = \frac{1}{(\sum_{j=1}^{c} N_j) - c} \sum_{j=1}^{c} \sum_{i=1}^{N_j} (\mathbf{x}_i^{(j)} - \hat{\mathbf{m}}^{(j)})(\mathbf{x}_i^{(j)} - \hat{\mathbf{m}}^{(j)})^T,$$

wobei c die Anzahl der Klassen, $\mathbf{x}_i^{(j)}$ das i-te Stichprobenelement, $\hat{\mathbf{m}}^{(j)}$ das arithmetische Mittel und N_j die Anzahl der Stichprobenelemente der Klasse $\omega = j$ bezeichnet.

b) Plotten Sie die Verteilungen (z.B. als Scatterplot) und die mittels LDA gefundene Entscheidungsgrenze.

c) Berechnen Sie sowohl den Trainingsfehler als auch den Testfehler (letzteren auf der Testmenge in *ldaTest.txt*).

4 Aufgabe UE-III.4

Wenn die N Beobachtungen Y_i iid und somit unkorreliert sind, ist die Varianz des Stichrobenmittels $\bar{Y}$ bekanntlich durch σ^2/N gegeben, wobei $\sigma^2 = Var[Y]$ die allen Beobachtungen zugrundeliegende Populationsvarianz bezeichnet.

Wir nehmen nun an, daß die Beobachtungen korreliert sind, d.h. für die Kovarianzmatrix der Beobachtungen $\mathbf{\Sigma} = (\sigma_{ij})_{1 \le i,j \le N}$ gelte

$$\sigma_{ij} = \begin{cases} \rho\sigma^2, \text{falls } i \neq j \\ \sigma^2, \text{falls } i = j \end{cases} . \tag{2}$$

Berechnen Sie mittels Lemma 2 (Vorwärtstransport der Varianz) die Varianz des Stichprobenmittels für korrelierte Beobachtungen. Was geschieht im Extremfall $\rho = 1$?