

1. Gau  -Klassifikator

Implementieren sie einen Gau  klassifikator, der zehn Klassen unterscheiden kann, und wenden Sie ihn auf den Datensatzes “Digits”¹ an. Dazu m  ssen sie die Trainings-Daten entsprechend den Ziffern separieren und f  r jede Klasse die Parameter μ und Σ berechnen.

Wenden Sie Ihren Klassifikator auf die Test-Daten an und erstellen Sie die Konfusionsmatrix.

Hinweise:

- (a) Die Dichtefunktion einer multivariaten Normalverteilung ist definiert als:

$$\mathcal{N}(\mu, \Sigma) = (2\pi)^{-\frac{k}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \mu)' \Sigma^{-1} (x - \mu)\right)$$

- (b) Vergleichen Sie den Logarithmus zweier Normal-Verteilungen, um einen Underflow zu vermeiden.

$$\mathcal{N}(\mu_1, \Sigma_1) > \mathcal{N}(\mu_2, \Sigma_2) \iff \log(\mathcal{N}(\mu_1, \Sigma_1)) > \log(\mathcal{N}(\mu_2, \Sigma_2))$$

- (c) Unter Umst  nden kann Σ nicht invertierbar sein. Nutzen sie in diesem Fall die Bibliotheksfunktion `pinv()` oder   hnliches. Alternativ k  nnen Sie, wie in der Vorlesung besprochen, Σ regularisieren. Statt Σ^{-1} berechnen Sie $(\alpha I + (1 - \alpha)\Sigma)^{-1}$ mit einem sehr kleinen α .

Hinweis: Bitte bearbeiten Sie die Aufgaben in Zweier-Gruppen und laden Sie alle Ergebnisse (Quelltext + Dokument mit Plots, Tabellen und Erl  uterungen) auf der Vorlesungsseite im Whiteboard hoch². Geben Sie die Namen beider   bungspartner an. Die Bewertung erfolgt bin  r (bestanden/nicht bestanden).

¹ <https://web.stanford.edu/~hastie/ElemStatLearn/data.html> → ZIP Code

² <https://kvv.imp.fu-berlin.de/portal/site/30550a4f-be9f-4be0-8ec2-a35a294ab3a7>