

Aufgabe 1 Fisher-Diskriminante

10 Punkte

Benutzen Sie den im Tutorium erläuterten Spam-Datensatz¹. Die Daten sind in einer CSV-Datei gegeben. Jede Zeile stellt Merkmale einer E-Mail dar. Die letzte Spalte gibt an, ob die entsprechende E-Mail als Spam angesehen wird. Eine Beschreibung der Merkmale finden Sie im gleichen Ordner. Teilen Sie den Datensatz in 80% Trainingsdaten und 20% Testdaten ein.

Implementieren Sie einen binären Klassifikator mittels der Fisher-Diskriminante. Kann man anhand der gefundenen Projektion beurteilen welche Merkmale am nützlichsten sind?

Geben Sie die Klassifikationsgenauigkeit an. Erstellen Sie die Konfusionsmatrix und beschreiben Sie diese kurz. Plotten Sie außerdem die projizierten Dichtefunktionen der Klassen.

¹<https://archive.ics.uci.edu/ml/datasets/spambase> → Data Folder bzw. Data Set Description