

Aufgabe 4.3

(1,5 Punkte)

Laden Sie sich das Jupyter-Notebook zu diesem Übungsblatt herunter und fügen Sie Ihre Featureberechnungen (siehe Aufgaben der 8. Woche) an der entsprechenden Stelle ein. Vervollständigen Sie die Methoden `normalize_data()` und `classify(train_data, test_data, cls_name)` so, wie im Notebook beschrieben.

Aufgabe 4.4

(1 Punkt)

Ein binärer Klassifizierer K_1 liefert auf den Validierungsdaten die nachfolgende Confusion Matrix.

		Ground Truth	
		Klasse A	Klasse B
Predicted	Klasse A	100	8
	Klasse B	2	5

Beantworten Sie die folgenden Fragen:

- (a) Wie viele Daten gehören zur Klasse A und wie viele zur Klasse B?
- (b) Berechnen Sie Precision, Recall, Accuracy und den F_1 -Score
- (c) Ist eine dieser Messungen ausreichend um die Qualität des Klassifizierers zu beschreiben? Wenn ja, welche Messung beschreibt die Qualität des Klassifizierers am besten. Falls nicht was ist das Problem?
- (d) Angenommen wir haben einen zweiten Klassifizierer K_2 der über eine Gleichverteilung zufällig entscheidet ob ein Sample zur Klasse A oder B gehört. Welche Accuracy hat K_2 ?
- (e) Welcher Klassifizierer ist für die korrekte Klassifizierung von Samples aus der Klasse B besser, K_1 oder K_2 ?
- (f) Was würden Sie empfehlen, um K_1 zu verbessern?

Falls Ihnen nicht alle Messungen bekannt sein sollten, dann sehen Sie hier nach³.

³https://en.wikipedia.org/wiki/Confusion_matrix

Aufgabe 4.5

(1,5 Punkte)

Erklären Sie die folgenden Klassifizierer in Ihren eigenen Worten (je ca 5-10 Sätze) mit Abbildungen und ggf. Berechnungsformeln:

- (a) k-Nearest Neighbor
- (b) Decision Tree
- (c) Support Vector Machine (SVM)
- (d) Multilayer Perceptron (MLP)