Künstliche Intelligenz WS 22/23

3. ÜBUNGSBLATT

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

Aufgabe 3.1 (4 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

Hierbei sei die Klasse A die positive Klasse und die Klasse B die negative Klasse.

Beantworten Sie die folgenden Fragen:

(a) Wie viele Daten gehören tatsächlich zur Klasse A und wie viele zur Klasse B?

(b) Berechnen Sie Precision, Recall, Accuracy und den F_1-Score

(c) Ist eine dieser Metriken ausreichend um die Qualität des Klassifizierers zu beschreiben? Wenn ja: Welche Metrik 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 Datenpunkt zur Klasse A oder B gehört. Welche Accuracy hat K_2?

(e) Wie können Sie ganz einfach einen Klassifizierer entwickeln, der eine Precision von 100% erreicht?

(f) Was würden Sie empfehlen, um K_1 zu verbessern?

Falls Ihnen nicht alle Metriken bekannt sein sollten, dann sehen Sie hier nach[3].

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

Aufgabe 3.2 (5 Punkte)

Bearbeiten Sie folgende Teilaufgaben:

a) Nennen Sie Vor- und Nachteile des k-Nearest-Neighbor (k-NN) Algorithmus.

b) Wie ändert sich die Ausgabe des k-NN, wenn wir für k große bzw. kleine Werte einsetzen?

c) Angenommen wir möchten mit dem k-NN Bilder klassifizieren. Nennen Sie eine Möglichkeit Bilder in eine Vektorform zu bringen.

d) Ist es sinnvoll den k-NN für Bildklassifizierung zu verwenden? Begründen Sie Ihre Antwort.

e) Welche Eigenschaften sollten Daten enthalten, damit sie gut durch den k-NN trennen lassen?

Aufgabe 3.3 (7 Punkte)

In dieser Aufgabe soll der k-Nearest-Neighbor Algorithmus implementiert werden. Sehen Sie sich dazu die Aufgaben im Jupyter-Notebook an und bearbeiten Sie diese.