

## Übungsblatt 2

Regression und Klassifikation mit linearen Modellen

Abgabe online auf ILIAS bis 14. November 2024, 12:00 Uhr

Ladet für die Abgabe bitte ein PDF mit der eingescannten Lösung für Aufgabe 1 sowie das bearbeitete Notebook für Aufgabe 2 hoch.

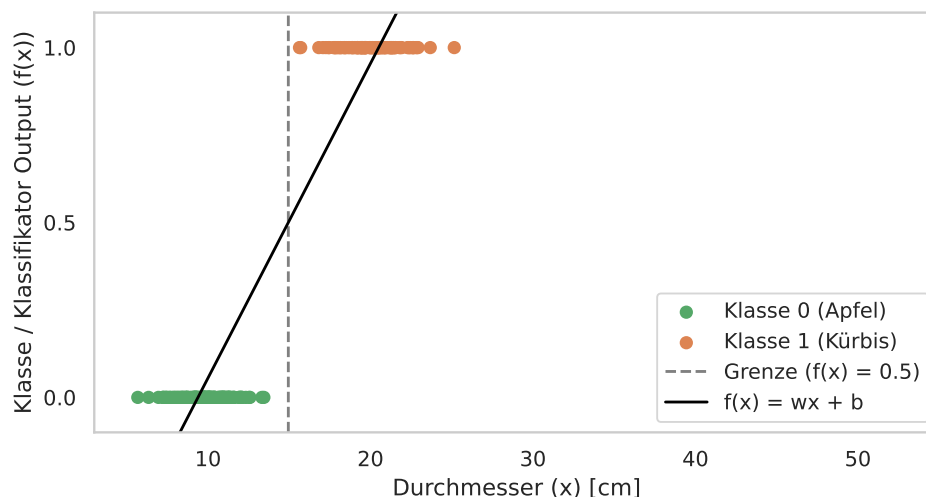
### Aufgabe 1 *Klassifikation*

(6 Punkte)

Wir möchten ein Modell zur Klassifizierung von Früchten entwickeln, wobei in unserem Datensatz jedes Stück durch einen  $i$ -dimensionalen Vektor  $x$  dargestellt ist. Wir beginnen mit zwei Fruchtsorten mit den Klassenlabels  $c_0 = 0$  (Apfel) und  $c_1 = 1$  (Kürbis). Wir wissen, dass der Datensatz linear separierbar ist und verwenden zunächst einen linearen Klassifikator der Form

$$f(x) = w^T x + b$$

wobei  $f(x) < 0,5$  der Klasse 0 und  $f(x) \geq 0,5$  der Klasse 1 entspricht. Dabei ist  $x$  der Eingangsvektor,  $w$  der Gewichtsvektor und  $b$  der Bias. Wir beginnen mit einem eindimensionalen Modell, das nur den Durchmesser der Früchte berücksichtigt, und erhalten den folgenden Klassifikator:



1. (2 Punkte) Erklärt, warum dieser lineare Klassifikator zusammen mit dem mittleren quadratischen Fehler ('mean square error', MSE) als Verlustfunktion nicht gut für Klassifikationsprobleme geeignet ist. Zeichnet zur Veranschaulichung eine modifizierte Version des gezeigten Plots, mit einem 1D linear separierbaren Datensatz, bei dem die Datenpunkte zweier Klassen mit der Grenze bei  $f(x) = 0.5$  nicht getrennt werden können.
2. (1 Punkt) Für einen eindimensionalen Datensatz können wir mit diesem Verfahren dennoch den Datensatz separieren, indem wir eine andere Grenze wählen. Begründet, weshalb das ab einem zweidimensionalen Datensatz nicht immer der Fall ist (*Hinweis: siehe Vorlesung 3 Slide 47 und 52*).
3. (2 Punkte) Erklärt, warum die Sigmoid-Funktion besser als Klassifikator geeignet ist. Welche Verlustfunktion wird für Klassifikationen anstelle von MSE verwendet und weshalb können wir die Ausgaben als Wahrscheinlichkeiten interpretieren?
4. (1 Punkt) Nachdem wir das Klassifikationsmodell trainiert haben, möchten wir bewerten, wie gut es die korrekten Klassen vorhersagen kann. Die Genauigkeit ('accuracy') ist die Metrik, die den Anteil der korrekt vorhergesagten Instanzen an der Gesamtzahl der Instanzen in einem Datensatz misst. Erklärt, warum und in welchen Fällen die Genauigkeit als Metrik dafür nicht ausreicht.

## **Aufgabe 2** *Notebook*

(14 Punkte)

Diese Aufgabe ist eine Programmieraufgabe. Bearbeitet das Notebook wie unter Übungsbetrieb im Vorlesungsinhalt Verzeichnis auf ILIAS beschrieben und folgt den Anweisungen dort.