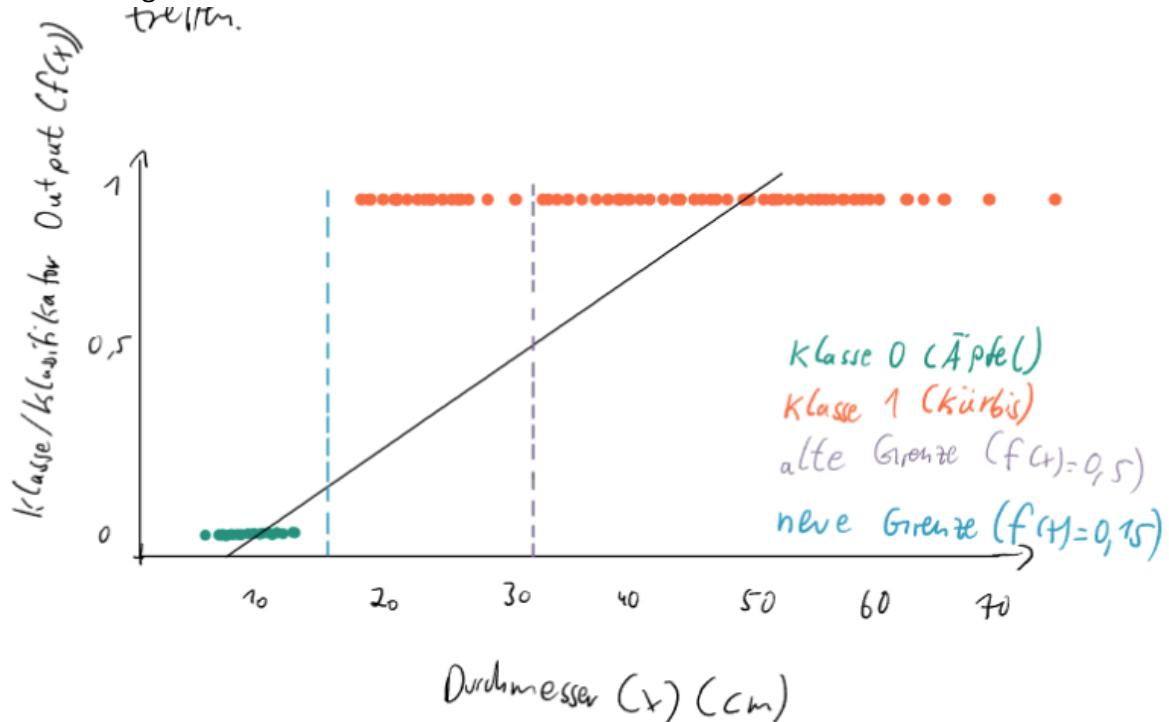


GKI – Übungsblatt 2 – Aufgabe 1

1.1

Beim MSE wird der Abstand zwischen dem vorhergesagtem und tatsächlichen Wert gemessen. Bei einer Klassifikation ist es aber wesentlich sinnvoller eine Einteilung nach richtig/falsch der Klassifizierung zu treffen.



Wenn wir nun im Beispiel den Datensatz um Riesenkürbisse erweitern, dann steigt der Durchschnitt des Durchmessers stark an. Dadurch verschiebt sich die Grenze von $f(x)=0,5$ welche Äpfel und Kürbisse trennt nach unten $\rightarrow f(x)=0,15$.

1.2

Sobald der Datensatz zweidimensional ist, kann es zur Durchmischung von Werten kommen. (Werte der einen Klasse werden durch Werte der anderen Klasse „umgeben“) Die Werte bilden keine homogenen Gruppen mehr. Dadurch ist es unmöglich mit einer Geraden die Daten klar zu unterteilen. Man benötigt (wesentlich) komplexere Abtrennungen, falls es überhaupt noch möglich ist.

1.3

Die Sigmoid Funktion ist besser geeignet, da sie die Aufgabe auf den Bereich $[0,1]$ beschränkt. Die Ausgabe kann, als die Unsicherheit der Klassifikation, als Wahrscheinlichkeit interpretiert werden. Ein Wert von z.B. 0,8 würde dann heißen, dass das Modell denkt das Objekt gehört in die Klasse 1 mit einer Unsicherheit von 20%.

1.4

Die Accuracy misst nur den Anteil korrekt identifizierter Datenpunkte an der Gesamtzahl von Datenpunkten. Bei einem großen Ungleichgewicht der Anzahl der beiden Klassen, kann eine hohe Accuracy bestehen, vor allem für die Häufig vorkommende Klasse. Gleichzeitig wird die wenig repräsentierte Klasse schlecht erkannt, da zu wenige Beispieldaten vorliegen.