

Aufgabensammlung

Aufgabe 1

Erläutern Sie den Unterschied zwischen Modell- und Hyperparametern.

Aufgabe 2

Erläutern Sie die Begriffe *Supervised* und *Unsupervised Learning*.

Aufgabe 3

Definieren Sie den Begriff *Feature Space*.

Aufgabe 4

Nennen Sie die Formel der linearen Regression mit Polynom-Trick vom Grad 2 für die Eingabe $x \in \mathbb{R}^2$.

Aufgabe 5

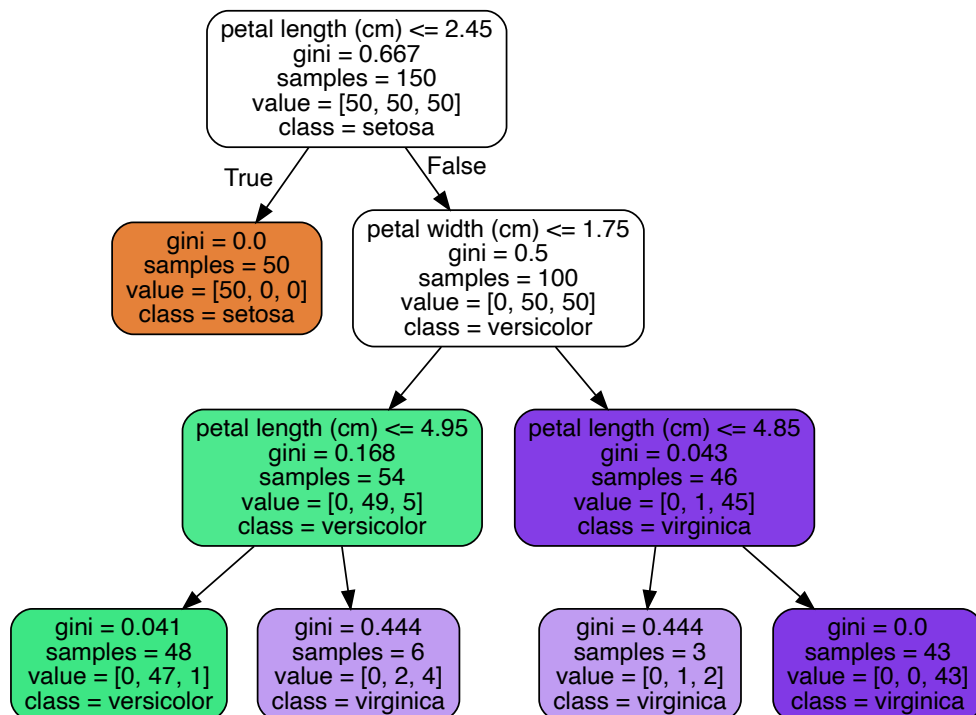
Beschreiben Sie den Unterschied zwischen *Hard-Margin*- und *Soft-Margin-Classifer* in Bezug auf eine *Support Vector Machine (SVM)*. Welche Konsequenzen hat die Wahl des einen oder anderen für ihre Machine-Learning-Anwendung?

Aufgabe 6

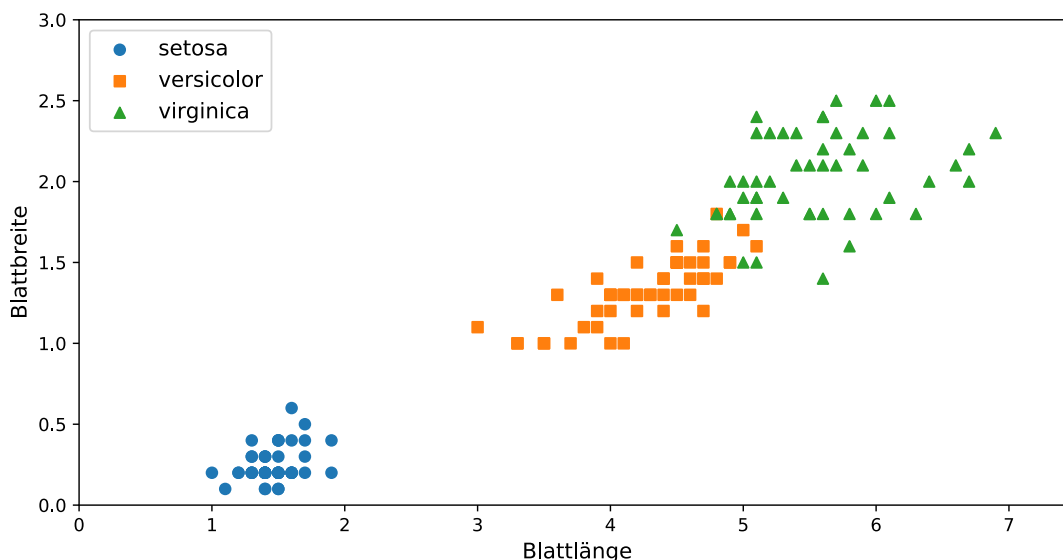
Nennen Sie die Formel zur Vorhersage eines Datenpunktes $x \in \mathbb{R}^2$ mittels einer *Support Vector Machine (SVM)*.

Aufgabe 7

Zeichnen Sie im *Feature Space* die *Entscheidungsgrenze* (Decision Boundary) des abgebildeten Entscheidungsbaums ein. Welcher Klasse wird der Datenpunkt mit den Merkmalen *petal length*=4.9 und *petal width*=1.73 zugewiesen?



Zeichnen Sie *hier* Ihre Lösung ein:



Aufgabe 8

Erläutern Sie die Schritte, die der Algorithmus *CART* (Classification and Regression Trees) durchläuft, um den nächsten Split eines *Decision Trees* zu finden.

Aufgabe 12

Gegeben seien die Ergebnisse einer Evaluierung verschiedener Machine Learning Modelle für die Prognose über eine Corona-Infektion. Für welches Modell würden Sie sich entscheiden? Begründen Sie Ihre Wahl.

Modell 1

Accuracy: 0,90

Patient (Zeile)/ Prediction (Spalte)	0	1
0	81	0
1	9	0

Modell 2

Accuracy: 0,80

Patient (Zeile)/ Prediction (Spalte)	0	1
0	67	14
1	4	5

Modell 3

Accuracy: 0,74

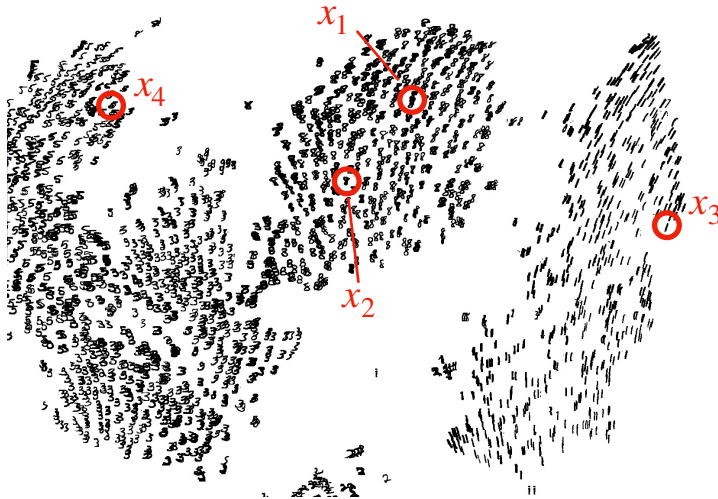
Patient (Zeile)/ Prediction (Spalte)	0	1
0	58	23
1	0	9

Aufgabe 13

Beschreiben Sie die Schritte, die zur Durchführung der *Principal Component Analysis (PCA)* notwendig sind, um die n Hauptkomponenten eines Datensets zu bestimmen.

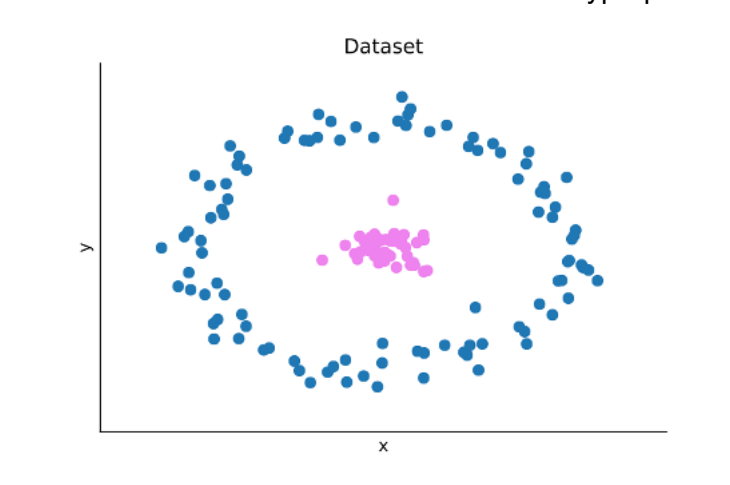
Aufgabe 14

Die Abbildung zeigt einen Ausschnitt aus dem MNIST-Datensatz, welches mithilfe des Algorithmus *t-Stochastic Neighbourhood Embedding* auf zwei Dimensionen reduziert wurde. Welche Aussage können Sie über die Datenpunkte x_1 und x_2 treffen? Welche Bedeutung hat der Abstand zwischen x_3 und x_4 ? Begründen Sie Ihre Aussage anhand der Funktionsweise des *t-SNE*-Algorithmus.



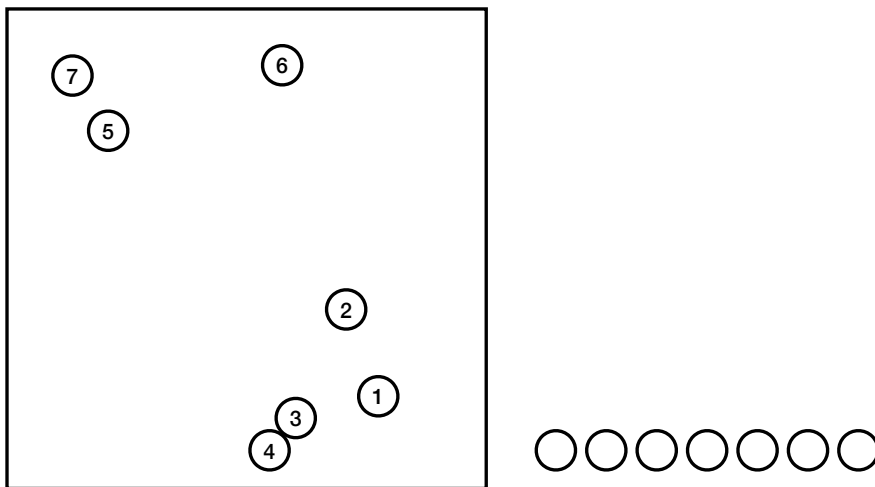
Aufgabe 15

Das abgebildete Datenset soll mithilfe des *k-Means*-Algorithmus gruppiert werden. Halten Sie diese Wahl für sinnvoll? Begründen Sie Ihre Antwort unter Einbeziehung der Funktionsweise von *k-Means* und verschiedener Werte für den Hyperparameter k .



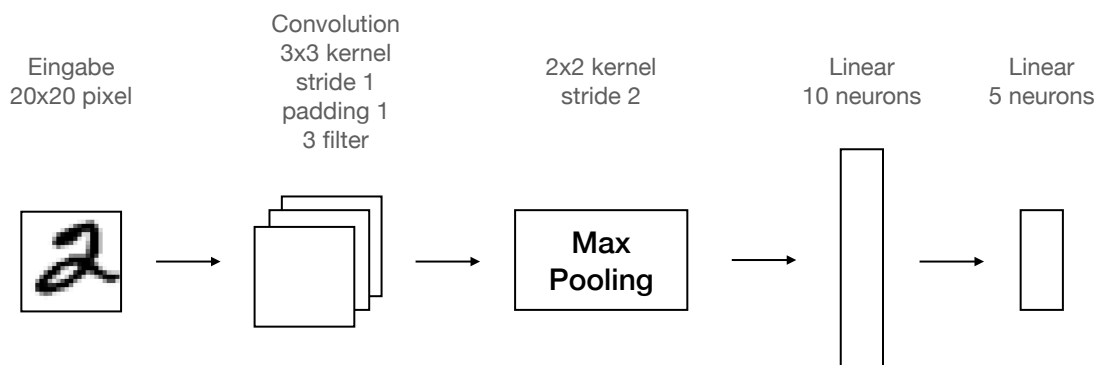
Aufgabe 16

Wenden Sie den Algorithmus *Agglomerative Clustering* mit dem Unähnlichkeitsmaß *Single Linkage* (mit euklidischer Distanz) auf den gegebenen Datensatz an und vervollständigen Sie das zugehörige Dendrogramm.



Aufgabe 17

Wieviele Modell-Parameter hat das abgebildete Neuronale Netz?



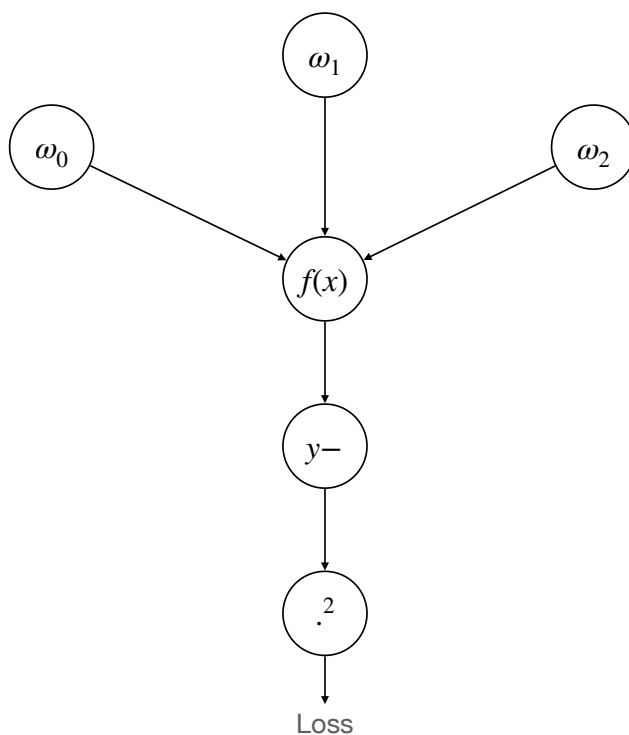
Aufgabe 18

Bestimmen Sie den Gradient $\frac{\partial L}{\partial \omega}$ der Kostenfunktion L an der Stelle x_i . Gehen Sie dabei analog zur *Backpropagation* vor und notieren Sie die Zwischenschritte im gegebenen Berechnungs-Graph.

$$x_i = (1, 3), \quad y_i = 2.3, \quad \omega = (0.1, 0.1, 0.1)$$

$$f(x) = \omega_0 x_1^2 + \omega_1 x_2^2 + \omega_2 x_1 x_2$$

$$L = \frac{1}{N} \sum_{i=1}^N (y - f(x))^2$$



$$\frac{\partial L}{\partial \omega_0} =$$

$$\frac{\partial L}{\partial \omega_1} =$$

$$\frac{\partial L}{\partial \omega_2} =$$

Aufgabe 19

Führen Sie eine *Convolution* auf dem Bild unter Anwendung des gegebenen *Kernels* durch. Gehen Sie dabei von einem *Zero-Padding* von 1 entlang jeder Dimension und einer Schrittweite (*stride*) von ebenfalls 1 in jede Richtung aus.

Kernel			Bild					Ergebnis				
0	-1	0	0	0	1	1	0					
-1	4	-1	0	1	1	1	0					
0	-1	0	1	1	1	1	0					
			0	0	1	1	0					
			0	0	1	1	0					

Aufgabe 20

Wenden Sie ein *2x2 Max-Pooling* auf das gegebene Bild mit einer Schrittweite (*stride*) von 2 in jede Richtung an.

7	3	8	0
5	5	9	0
5	5	1	10
3	4	0	0
