MachineLearning for Visual Computing
Aufgabenblock 1

Christian Brändle, Doris Antensteiner

November 25, 2014

Abstract

0.1 Einfaches Perceptron - Datengeneration

Gegeben sind vier Datensets a 100 Beobachtungen von 2-dimensionalen Eingangsdaten, welche normalverteilt sind. Die Figuren in Tabelle 1 verdeutlichen dies.

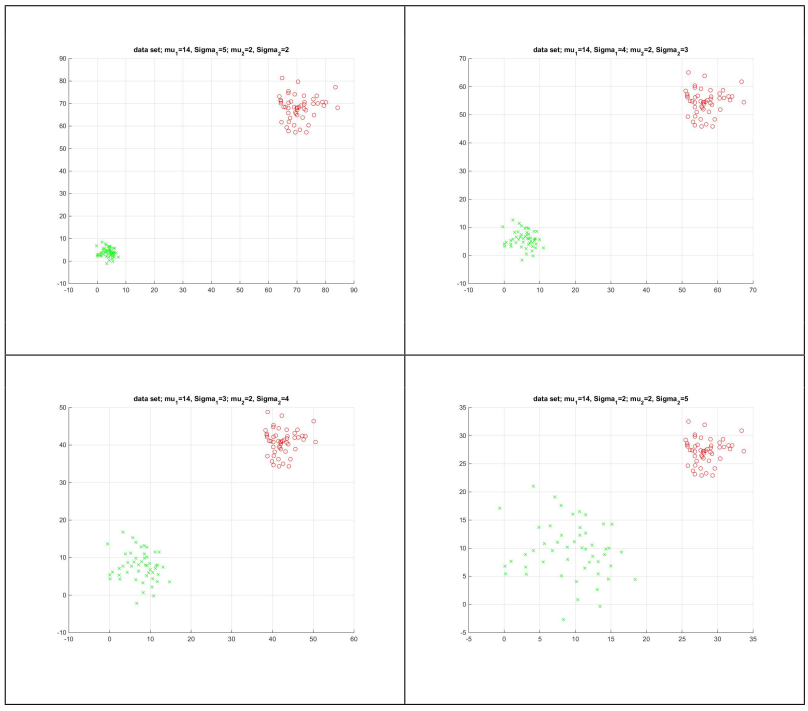


Table 1: Iteration distribution over gamma

0.2 Einfaches Perceptron - Perceptrontraining

0.2.1 Anzahl Iterationen

Die Anzahl der Iterationen hängt maßgeblich von der *margin* der separierbaren Daten ab. Je kleiner die margin, desto mehr Iterationen sind notwendig, bis der Algorithmus terminiert.

0.2.2 Welchen Einfluß hat die Schrittweite

Der Einfluß der Schrittweite hängt sowohl von den Eingangsdaten als auch vom gewählten Algorithmus ab. Es ist zu beobachten, daß der batch-Algorithmus keine großen Abweichungen zeigt, egal welches γ gewählt wird. Beim online-Verfahren ist der Einfluß größer, das heißt die Varianz in den ermittelten It-

erationen ist höher, aber auch hier ist kein eindeutiger Bereich über alle Simulationen auszumachen, wo ein gegebenes γ die Iterationen des Algorithmus wesentlich reduziert. Zu sehen ist die Verteilung der Iterationen über ein γ von 0.1 bis 4 für die jeweiligen unterschiedlichen Datensätze in Tabelle 2.

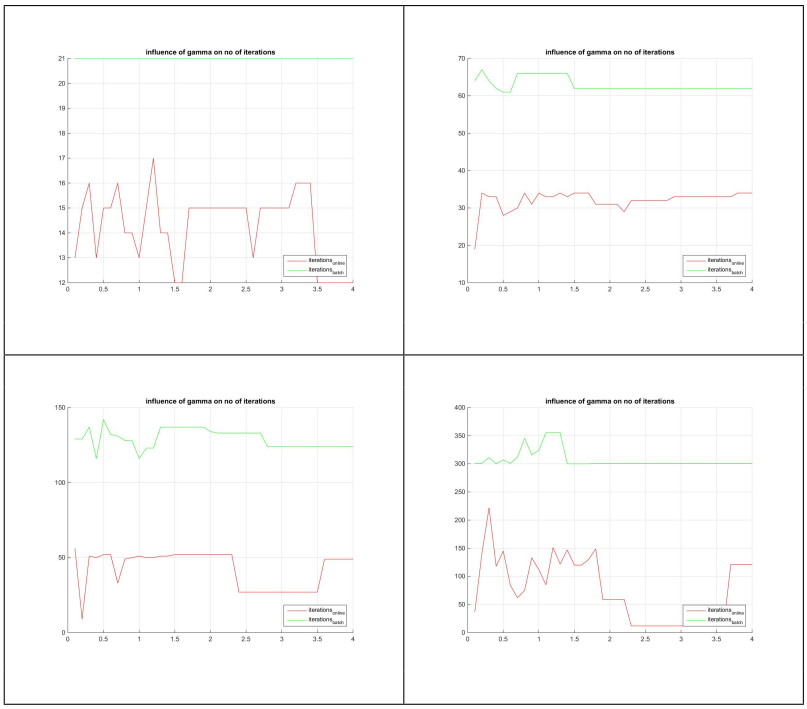


Table 2: Iteration distribution over gamma

0.2.3 Daten und Entscheidungsgrenzen im $\mathbb{R}^2$

Die ermittelten Entscheidungsgrenzen des batch-learning sowie des online-learning Algorithmus sind für die einzelnen Datensätze und ein gegebenes γ von 1 in Tabelle 3 zu sehen.

0.2.4 Wie ist das Verhalten bei nicht linear separierbaren Daten

Der Algorithmus terminiert nicht da bei jedem Durchlauf Daten gefunden werden, welche in der falschen Klasse landen.

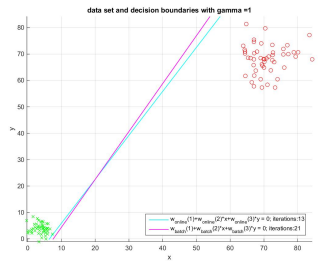
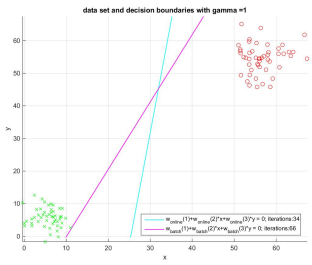
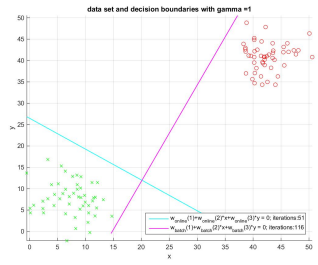
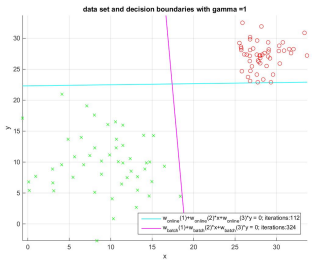
	
set_1: $\mu_1 = 10$, $\Sigma_1 = 5$ set_2: $\mu_2 = 2$, $\Sigma_2 = 2$	set_1: $\mu_1 = 10$, $\Sigma_1 = 4$ set_2: $\mu_2 = 2$, $\Sigma_2 = 3$
	
set_1: $\mu_1 = 10$, $\Sigma_1 = 3$ set_2: $\mu_2 = 2$, $\Sigma_2 = 4$	set_1: $\mu_1 = 10$, $\Sigma_1 = 2$ set_2: $\mu_2 = 2$, $\Sigma_2 = 5$

Table 3: Iteration distribution over gamma

0.3 Lineare Regression

Wir erstellen Daten im Bereich von $[0, 5]$ mit einer Schrittweite von 0.1 aus der Funkion:

$$y = x^2 - Gx + 1 \tag{1}$$

Des weiteren erstellen wir eine Trainingsmenge aus jedem sechsten Datenpunkt und fügen je einen Zufallswert aus $N(\mu = 0, \Sigma = 0.7)$ hinzu.

0.3.1 Gewichtsvektor mittels Gradientenabstieg bei quadratischer Fehlerfunktion

0.3.2 Optimaler Gewichtsvektor w^*

Den optimalen Gewichtsvektor w^* kann man über die Bestimmung der Pseudoinversen A^+ für die Gleichung $Aw^* = b$ ermitteln. Die Gleichung $w^* = A^+b$

bestimmt das optimale w^*. Dabei berechnet sich A^+ wie folgt:

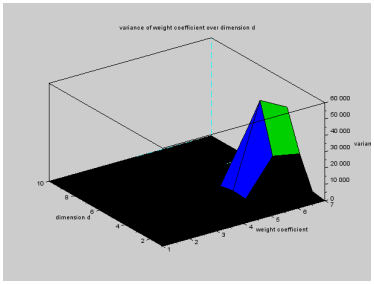
$$A^+ = (AA^T)^{-1} \quad \text{für invertierbare } AA^T \tag{2}$$
$$A^+ = (AA^T)^{-1} + \lambda I \quad \text{für nicht invertierbare} AA^T \text{ mit } \lambda << 1 \tag{3}$$

Das ermittelte Ergebnis für w^* mit normalisierten Daten aus dem $\mathbb{R}^2$ lautet: $w^* = (1.0106678, -10.0086, 1.8960336, 0.0262634)$.

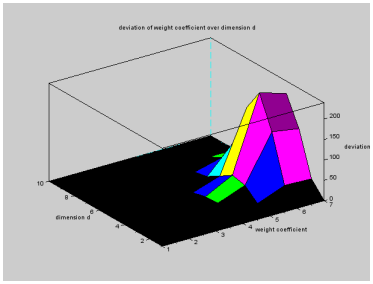
0.3.3 Konvergenzverhalten in Abhängigkeit von γ

0.3.4 Erwartungswert und Varianz von w^* im Bezug zu Dimensionalitt d von $f(x)$

Erwartungswert und Varianz von w^* im Bezug zu Dimensionalitt d von $f(x)$ zeigt, dass je höher die Dimensionalitt d steigt, desto höher auch die Varianz *Sigma* der ermittelten Werte für w^* für verrauschte Eingangsdaten wird. Dies ist auch klar, da hherdimensionale Koeffizienten von w^* höhere Frequenzen repräsentieren, welche durch das Verrauschen der Eingangsdaten stärker betroffen sind als niedrigere Frequenzen. Interessanterweise steigt die Varianz aller Koeffizienten von w^* in Tabelle 4 nur bis zu einer Dimensionalität des Problems von 6, danach fällt sie wieder. Es ist aber anzunehmen, dass dies zum Einen aufgrund von Unzulänglichkeiten beim generieren der Pseudoinversen zur Berechnung von w^* liegt und zum Anderen von der fehlenden Betrachunng noch höherer Dimensionen, wo sich globale Anstieg der Varianz sicherlich fortsetzt.



variance of weight coeffinients



deviation of weight coeffinients

Table 4: variance and deviantion of w^* over dimension

Des weiteren ist festzustellen, dass für höherdimensionale Komponenten für w^* durch overfitting die Varianz zunimmt.

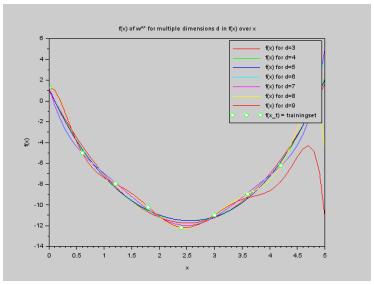


Figure 1: overfitting of y^* curves over dimension