

6. Übungsblatt

Aufgabe 28 Risiko und empirisches Risiko

- a) Sie haben einen blauen und einen roten Würfel und sind an zwei Auskommen interessiert:
- A : das Ergebnis des roten Würfels
 - B : die Summe der beiden Würfel

Berechnen Sie die Wahrscheinlichkeiten $P(B > 6 | A = k), k \in \{1, 2, 3, 4, 5, 6\}$

- b) Nun wollen wir die Wahrscheinlichkeiten auf eine andere Weise bestimmen. Dafür schätzen Sie die bedingten Wahrscheinlichkeiten aus a) nur aus den folgenden Würfeln:

rot	blau
3	4
1	6
4	1
2	6
6	2
2	5
6	1
5	2
3	5
4	2
5	1

- c) Setzen Sie die Ergebnisse aus a) und b) ins Verhältnis zu dem was bei der Optimierung Neuronaler Netze passiert.

Aufgabe 29 Fehlerfunktionen und wie sie zu finden sind

Beim Trainieren von neuronalen Netzen können unterschiedliche Fehlerfunktionen verwendet werden. Dabei werden in der Regel Parameter θ gesucht die $J(\theta) = \mathbb{E}_{x,y \sim p} \log(P(y|x; \theta))$ maximieren. Also den Erwartungswert, dass der richtige Wert y eine hohe Wahrscheinlichkeit gegeben dem zugehörigen Wert x hat. Dieses Verfahren nennt sich Maximum-Likelihood Schätzer. In einer Regressionsaufgabe ist $P(y|x; \theta) = \mathcal{N}(y, f(x; \theta), 1)$, wobei

$$\mathcal{N}(y, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right)$$

Das bedeutet, dass die Wahrscheinlichkeit hoch ist, wenn der Zielwert nah an dem vorhergesagten Wert ist

Ermitteln Sie die zu optimierende Funktion, wenn wir gegebenes $P(y|x; \theta)$ nehmen und $J(\theta)$ maximieren wollen (Einsetzen und vereinfachen). Erklären Sie, wie sich die Fehlerfunktion verhält, wenn der Fehler besonders groß oder besonders klein wird

Aufgabe 30 Fehlerfunktion: Kreuzentropie

Gegeben sei ein Datensatz von Bildern von fünf verschiedenen Tieren. Jedes Bild zeigt genau ein Tier. Jedes Bild hat zudem ein Label P , welches das abgebildete Tier als One-Hot-Kodierung wie folgt angibt:

Tier	Label
Hund	[1 0 0 0 0]
Katze	[0 1 0 0 0]
Pferd	[0 0 1 0 0]
Adler	[0 0 0 1 0]
Kuh	[0 0 0 0 1]

- Wofür ist die Entropie bzw. die Kreuzentropie im allgemeinen ein Maß? Warum und wie können wir dies für Neuronale Netze nutzen?
- Nehmen Sie an ein Netz, das auf diesen Daten trainiert, generiert für das Bild eines Hundes die Ausgabe $Q_1 = [0.45 \ 0.25 \ 0.03 \ 0.07 \ 0.2]$. Berechnen Sie die Kreuzentropie $H(P, Q_1)$. Das Netz klassifiziert das Bild korrekt als Hund. Warum ist die Ausgabe trotzdem problematisch?
- Das Netz hat nun eine längere Zeit trainiert und generiert für das selbe Bild die Ausgabe $Q_2 = [0.97 \ 0.01 \ 0.003 \ 0.007 \ 0.01]$. Berechnen Sie erneut die Kreuzentropie $H(P, Q_2)$. Was können wir basierend auf diesen Werten über unser Netz aussagen?

Aufgabe 31 Aktivierungsfunktion: Softmax

Im überwachten Lernen ist ein häufiger Anwendungsfall die Klassifikation. Ein typisches Beispiel ist die Klassifikation von Bildern wie beispielsweise in der ILSVRC vom <http://www.image-net.org/>.

Der quadratische Fehler ist für Klassifikationsaufgaben nicht geeignet. Aus diesem Grund verwendet man in der die Kreuzentropie : $H(P, Q) = - \sum_{i=1}^n P(i) \log Q(i)$, wobei P und Q jeweils Wahrscheinlichkeitsverteilungen sind. Das bedeutet, dass unsere Netzwerke eine Wahrscheinlichkeitsverteilungen als Ausgabe haben muss. Nehmen wir eine lineare Ausgabe z aus dem Netzwerk an. Auf diese Ausgabe wird dann die Softmaxfunktion angewandt:

$$\text{softmax}(z)_i = \frac{\exp(z_i)}{\sum_j \exp(z_j)}$$

- Nehmen Sie an, dass Bilder von fünf verschiedenen Tieren als Datensatz genutzt werden. Die Labels sind dabei wie folgt one-hot Codiert :

Tier	Label
Hund	[1 0 0 0 0]
Katze	[0 1 0 0 0]
Pferd	[0 0 1 0 0]
Adler	[0 0 0 1 0]
Kuh	[0 0 0 0 1]

Die Ausgabe des Netzes sei wie folgt:

Eingabebild	Ausgabe y_i
Katze	[6 5 2 3 1]
Pferd	[4 3 8 1 -2]
Pferd	[4 4 1 -1 0]

Berechnen Sie die Softmaxausgaben.

- Da wir bei Klassifikation die Kreuzentropie minimieren wollen $H(labels, prediction)$. Berechnen Sie die Kreuzentropie für die Vorhersagen aus der Tabelle.
- Wie Sie bei der Berechnung sehen, beinhaltet die Kreuzentropie den Term $-\log(\text{softmax}(z)_i)$. Setzen Sie die Softmaxfunktion ein und vereinfachen Sie so weit wie möglich. Argumentieren Sie warum dieser Term gut für die Optimierung geeignet ist.

Aufgabe 32 Bonus: Automatische Differenzierungsframework (4)

In dieser Aufgabe wollen wir Optimierungs und Regularisierungsverfahren aus der Vorlesung implementieren. Es gibt erneut eine Vorlage für diese Aufgabe.

- a) Als erstes wollen wir zwei neue Optimierungsstrategien implementieren. Dafür implementieren Sie die Initialisierung und den Schritt für den Optimizer mit Manhattantraining und für das Gradientenverfahren mit Momentum.
- b) Als nächstes implementieren Sie die Normregularisierung mit L_1 und L_2 . Dafür sollten Sie zunächst die *abs* Operation in unserer Tensorklasse implementieren.
- c) Als letztes wollen wir mit unseren Optimizern und Regularisierungen experimentieren. Orientieren Sie sich dafür an folgenden Schritten:
 - Laden Sie den Datensatz herunter (<http://archive.ics.uci.edu/ml/datasets/communities+and+crime>) und legen Sie ihn in den Ordner zum jupyter-notebook.
 - In der Vorlage sind bereits die Funktionen zum laden der Daten bis zum Anfang der Trainingsschleifen.
 - Wählen Sie eine Architektur und implementieren Sie den Aufruf.
 - Trainieren Sie ihr Model mit unterschiedlich kombinationen von Optimizern und Regularisierungen und dokumentieren Sie welche Verhalten Sie entdecken. (Mindestens 3 verschiedene Kombinationen) Achten Sie dabei auf die Testperformance und Konvergenzdauer.