Prof. Dr.-Ing. Sebastian Stober, Michelle I. Bieber, Jan-Ole Perschewski, Christian Wustrau

# 8. Übungsblatt

## Aufgabe 32 Visualisierung von Gewichten

Betrachten Sie folgende Funktion $f(x) = relu(xW)$, wobei $x = (x_1, x_2, x_3, x_4, x_5)$ und

$$W = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

Initialisieren Sie $x = 0$ und maximieren Sie den zweiten Spaltenvektor $(xW)_1$ mit Gradientenaufstieg für zwei Schritte mit einer Lernrate von 1.

- Interpretieren Sie, was das Ergebnis des Gradientenaufstiegs zeigt.
- Betrachten Sie die Berechnung von $xW$. Überlegen Sie, bei welcher Architektur Sie diese Art von Berechnung bereits gesehen haben.

## Aufgabe 33 Fehlerfunktionen und wie sie zu finden sind

Beim Trainieren von neuronalen Netzen können unterschiedliche Fehlerfunktionen verwendet werden. Dabei werden in der Regel Parameter $\theta$ 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 $\mathcal{N}(y, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(y-\mu)^2}{2\sigma^2})$. 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.
- Erklären Sie, wie sich die Fehlerfunktion verhält, wenn der Fehler besonders groß oder besonders klein wird

## Aufgabe 34 Fehlerfunktion: Kreuzentropie

Gegeben sei ein Datensatz vn 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 Q1 = [0.45 0.25 0.03 0.07 0.2].
  Berechnen Sie die Kreuzentropie H(P,Q1). 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 Q2 = [0.97 0.1 0.03 0.07 0.1].
  Berechnen Sie erneut die Kreuzentropie H(P,Q2). Was können wir basierend auf diesen Werten über unser Netz aussagen?

## Aufgabe 35 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/. Um diese Aufgabe zu lösen eignenen sich unsere bisherigen Fehlerfunktionen nicht. Dafür wird erneut eine Maximum-Likelihood Schätzung verwendet: $\max_\theta \mathbb{E}_{x,c\sim p} \log(P(y = c|x;\theta)) \Leftrightarrow \min_\theta \mathbb{E}_{x,c\sim p} - \log(P(y = c|x;\theta))$

Für Klassifikationsprobleme wird in der Regel die Softmaxfunktion als $f_{act}$ verwendet.

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

- Nehmen Sie an, dass wie in Aufgabe 34 Bilder von fünf verschiedenen Tieren als Datensatz genutzt werden. 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 den Softmax der Ausgaben. Warum wird der Softmax der Ausgabe und nicht die direkte Ausgaben zur Klassifikation eingesetzt?

- Softmax wird oft in Verbindung mit der negativen Log-Likelihood Funktion $L(y) = \mathbb{E}_{x,c\sim p} - \log(P(y = c|x;\theta))$ als Fehlerfunktion verwendet.
  Berechnen Sie $-\log(softmax(z)_i)$. Welche Eigenschaft sorgt dafür, dass der Fehler einen hilfreichen Gradienten liefern kann ?