Exercise Sheet	02	**1. Team Partner:**	Lennart Slusny
Task	01	**2. Team Partner:**	Joschka Strüber

(a) How can LSTMs and other gated memory units learn to count and solve other discrete tasks? What is expected to happen during training?

Answer:

- time steps of calculations act as natural source of discreteness

wie läuft
ing ab? W
wird verg
was durch
lassen, w
funktionie
BPPTT?

(b) Consider a single LSTM unit with D memory cells. Starting with

$$\delta_\phi^t = \frac{\partial E}{\partial \operatorname{net}_\phi^t} \tag{1}$$

derive the forget gate gradient of the LSTM:

$$\delta_\phi^t = \varphi_\phi'(\operatorname{net}_\phi^t) \sum_{c=1}^{D} \zeta_c^t s_t^{t-1} \tag{2}$$

Answer:

V. 2 F. 23,
F. 21+23

(c) Can bidirectional RNNs be effectively used for online classification, that is generating a classification label for every new input immediately? Explain briefly.

Answer: No, that is in general not possible. To make inference, a bidirectional RNN needs both pretemporal and posttemporal context to compute an output. However, in online classification, only the pretemporal context is available. To get the posttemporal context, we would have to get to the end of a timeseries and compute the recurrence back in time, which is impossible in this context.

Task	02

(a) Name three forms of commonly applied regularization techniques for neural networks and describe their effects.

Answer:

Early Stop
Weight De
Label Sm
ing (V. 2 F

(b) Explain the internal covariate shift and how it has been addressed in the literature.

Answer:

V. 2, F. 49
Output je
Layer hän
von Statis
tiken der
puts ab, d
verändern
mit jeden
Trainingss
chritt auf
veränderl
Gewichte

(c) Consider the following statement: The more parameters a neural network has the better it will generalize. Is this statement true?

Answer:

(d) What is the effect of residual blocks on the gradient flow in deep networks?

Answer:

V 2, F. 61:
Backward
Gradient
entlang R
ual Conne
tion addi
propagier

Task 03

(a) What is the general purpose of autoencoders (AEs) and what are their main components? Give an illustration.

Answer:

V3 F4: Co
press inst
and gene
similar or
encoder a
decoder s
ture

(b) Briefly explain the purpose of the KL divergence.

Answer:

Misst die
Unähn-
lichkeit e
Verteilung
von einer
erenzvert
Q. Misst d
erwartete
zusätzlich
Shannon-
Informati
wenn Q
anstelle
der wahr
Verteilung
P verwen
wird.

(c) Given two discrete probability distributions Q, P. Show that $\mathrm{KL}(Q||P) = \mathrm{KL}(P||Q)$ does not hold in general.

Answer:

Super sin
pel mit p
$0.3, p_2 =$
$p_1 = 0.7,$
$q_2 = 0.5$

(d) Briefly describe the re-parametrization trick and explain why it is necessary to train a VAE.

Answer:

(e) Name a frequently mentioned drawback of VAEs and give one example from the literature that addresses this issue.

Answer:

(f) What are the major differences of the variational autoencoders (VAEs) in comparison with standard autoencoders? Relate your answer to the two sub-losses that are optimized during VAE training.

Answer:

(g) Bonus: Show that the KL-divergence between two Gaussian distributions simplifies in the following form, if the reference distribution is the standard normal:

$$\mathrm{KL}(N(\mu, \mathbf{\Sigma})||N(\mathbf{0}, \mathbf{I})) \tag{3}$$

V3 F12: V
wende
N(0,1) als
Zielvertei
das sorgt
für simpl
Form der
Divergenc
differenzi
ist.

Task 03

(a) Illustrate the major structure of GANs and name the essential components. Briefly describe the principle.

Answer:

Generato
generiert Daten
basierend
random i
Discrimin
muss gen
ierte und
Daten un
scheiden.

(b) Elaborate on why the training of GANs is considered as particularly difficult/challenging and how it is addressed. Comment on what happens if the discriminator dominates the generator and vice versa.

Answer:

Joint opt
tion von
verschied
Netzwerk
Trainings
beider Ne
müssen
anspruch
aber nich
möglich s
Discrimin
tor domin
generator
erator ler
nicht wie
Daten ger
ieren kan
die Discri
nator übe
ten. Gen
tor domin
Discrimin
tor: Disc
inator üb
keinen Dr
auf Gene
aus besse
Daten zu
gen und
stagniert.