# Exercise 22

Mittwoch, 29. Juni 2022 15:40

a) A loss function describes the cost or loss of precision/information when performing a prediction. It gives bigger values if the "distance" of the prediction from the real value is greater. Since the prediction depends on the parameters used in the classifier/predictor the loss function can be used to estimate good parameters by minimising the loss function.

b) The loss function can be minimised by looking for local minimum by variating the parameters.

Practically if you have a loss function that directly relies on the parameters you can follow its gradient to a local minimum while adjusting the parameters towards the descending gradient in each step.

c)

The purpose of activation functions is to add non-linearity to a neural network. It allows to use many neurons that perform linear transformations $x' = Wx + b$ with different weights. Without an activation function the layers of nodes would collapse to 1 single transformation because of the linearity.

d)

A neuron is a node in a neural network.
Similar to the neurons in the human brain it receives and processes data. A neuron can receive data from other neurons in layers behind it and applies the linear transformation (weights) and passes the weighted data to other neurons depending on the activation function.

e)

1) Text, image & voice recognition
2) Creating images/text

**2)** Creating images / text

**3)** Artifical Intelligence in Games

- Big amounts of data need to be handled in this examples also there is a lot of easy accessible training data
- These tasks are very complex to solve and good features can't be easily generated.

The neural network can handle big amounts of data and can also solve complex classification tasks through its many free parameters and neural layers.

# Exercise 23

**a)**

$\dim[x_i] = M$ (column vector)

$\dim[C] = 1$ (scalar)

$\dim[W] = M \times M$

$\dim[b] = M$ (column vector)

$\dim[\nabla_W C]$

$\dim[\nabla_{f_i} C]$

$\dim\left[\frac{\partial f_{k,i}}{\partial W}\right]$

$\dim\left[\frac{\partial f_{k,i}}{\partial b}\right]$

$$\nabla_{f_a} C(f) = \frac{1}{m} \sum_{i=1}^{m} \left[ \frac{\partial}{\partial f_a} \left( - \sum_{k=1}^{K} \underbrace{\mathbb{1}(y_i = k)}_{\substack{=0 \text{ for} \\ k \neq a}} \log \frac{\exp(f_{k,i})}{\sum_j \exp(f_{j,i})} \right) \right]$$

$$\Rightarrow \nabla_{f_a} C(f) = -\frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial f_a} \mathbb{1}(y_i = a) \log \left( \frac{\exp(f_{a,i})}{\sum_j \exp(f_{j,i})} \right)$$

$$\frac{\partial}{\partial f_a} \mathbb{1}(y_i = a) \log \frac{\exp(f_{a,i})}{\sum_j \exp(f_{j,i})} = \mathbb{1}(y_i = a) \frac{\sum_j \exp(f_{j,i})}{\exp(f_{a,i})} \cdot \frac{\partial}{\partial f_a} \frac{\exp(f_{a,i})}{\sum_j \exp(f_{j,i})}$$

$$= \mathbb{1}(y_i = a) \cancel{\frac{\sum_j \exp(f_{j,i})}{\exp(f_{a,i})}} \cdot \frac{\sum_j \exp(f_{j,i}) \cdot \cancel{\exp(f_{a,i})} - \exp(f_{a,i})^{\cancel{2}}}{(\sum \exp(f_{j,i}))^{\cancel{2}}}$$

$$= \mathbb{1}(y_i = a)\, \cancel{\frac{\sum_j \exp(f_{j,i})}{\exp(f_{a,i})}} \cdot \frac{\sum_j \exp(f_{j,i}) \cdot \cancel{\exp(f_{a,i})} - \exp(f_{a,i})^{\cancel{2}}}{\left(\sum_j \exp(f_{j,i})\right)^{\cancel{2}}}$$

$$= \mathbb{1}(y_i = a)\, \frac{\sum_j \exp(f_{j,i}) - \exp(f_{a,i})}{\sum_j \exp(f_{j,i})}$$

$$= \mathbb{1}(y_i = a) \cdot \left(1 - \frac{\exp(f_{a,i})}{\sum_j \exp(f_{j,i})}\right)$$

$$\Rightarrow \nabla_{f_a} C(f) = \frac{1}{m} \sum_{i=1}^{m} \mathbb{1}(y_i = a) \left( \frac{\exp(f_{a,i})}{\sum_j \exp(f_{j,i})} - 1 \right)$$