Practical 3.4

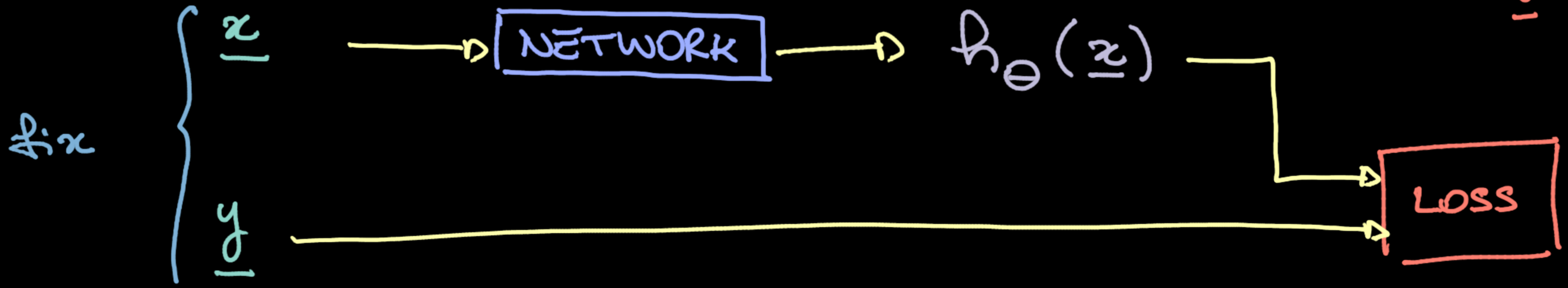
Convolutional Neural Networks – Loss functions

Overview

- Loss function
- Mean squared errors loss
- Regression and classification
- Sigmoid + cross entropy loss
- Softmax + negative log-likelihood loss
- Triplet embedding loss
  - Motivation
  - Theory
  - Implementation

Loss and cost function

$$\mathcal{L}(h_\theta(\underline{x}), \underline{y}) \in \mathbb{R}$$



$$J(\theta) \qquad \frac{\partial J(\theta)}{\partial \theta} \qquad \theta \to \theta - \eta \frac{\partial J}{\partial \theta}$$

$$(\quad)^T$$ for Jacobian notation

Mean squared error (MSE)

$(*) \quad + \frac{\lambda}{2m} \sum_{j \neq 0}^{3\times} \left(\Theta_{ij}^{(\ell)} \right)^2$

Regularisation (to fight overfitting)

$$J(\Theta) = \frac{1}{2m} \sum_{i=1}^{m} E_i^2 = \frac{1}{2m} \sum_{i=1}^{m} \| h_\Theta(\underline{x}^{(i)}) - \underline{y}^{(i)} \|^2 =$$

$$= \frac{1}{2m} \sum_{i=1}^{m} (h_\Theta(\underline{x}) - \underline{y})^T (h_\Theta(\underline{x}) - \underline{y}) \quad (*)$$

- Bounding boxes pred. $h_\Theta(\underline{x}) \in \mathbb{R}^4$

- Steering angle $h_\Theta(\underline{x}) \in \mathbb{R}$

$\underline{z}^{(L)}$: logits , $z_k^{(L)}$: logit

For regression

$h_\Theta(\underline{x}) = \underline{z}^{(L)}$

For classification

$h_\Theta(\underline{x}) = \underline{a}^{(L)} = \sigma(\underline{z}^{(L)}) \in [0,1]^k$

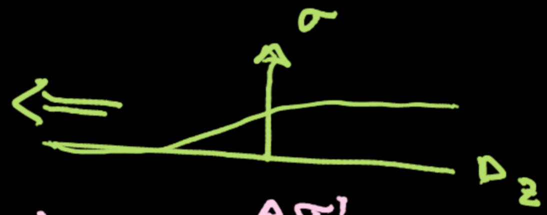
$a_k^{(L)}$: probability of class k

$\underline{y} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \leftarrow k=4$

Regression and logistic regression

$$\underline{x} \quad \begin{cases} y = 1 \\ h(\underline{x}) = 1 \end{cases} \qquad E = \frac{1}{2}\, 0 = 0$$

$$\underbrace{h(\underline{x}) = 0}_{\to\ 1/2} \qquad E = \frac{1}{2}\,(1)^2 = \frac{1}{2}$$



Using MSE with $\sigma(\cdot)$ is terribly slow => CROSS ENTROPY loss function

Cross entropy (I)

$y = 1, \quad h_\Theta(\underline{x}) \nearrow 1$ ok $\Rightarrow$ loss $= 0$

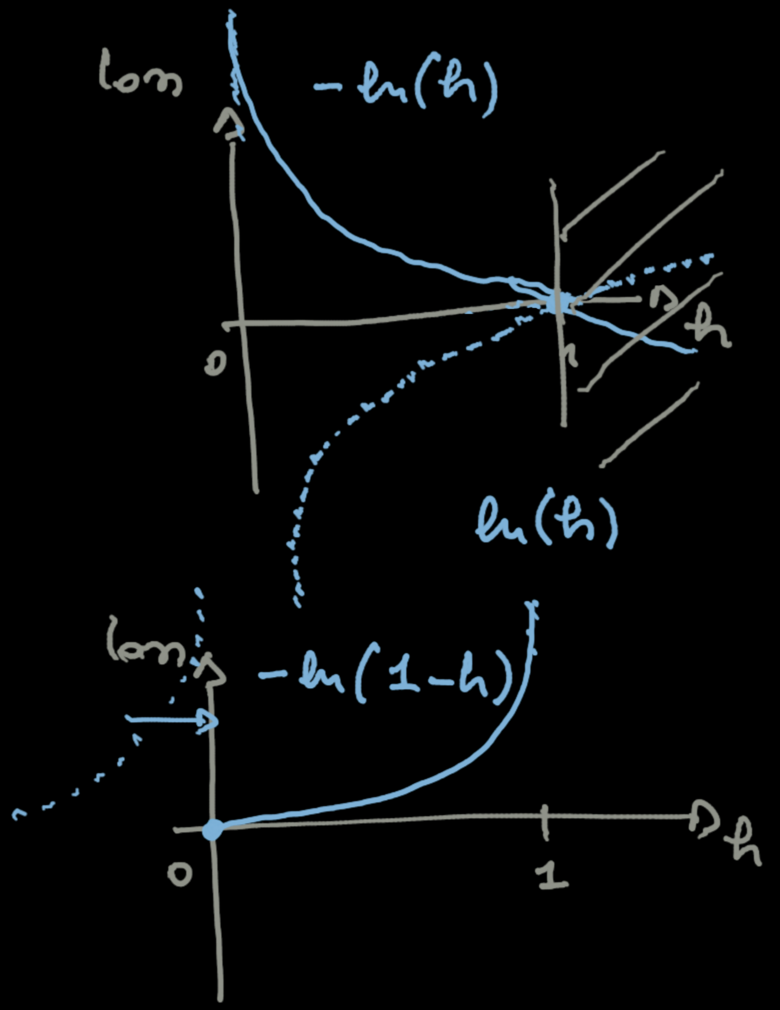
$\phantom{y = 1, \quad h_\Theta(\underline{x})} \searrow 0$ ko! $\Rightarrow$ loss $\to +\infty$

$y = 0, \quad h_\Theta(\underline{x}) \nearrow 0$ ok $\Rightarrow$ loss $= 0$

$\phantom{y = 0, \quad h_\Theta(\underline{x})} \searrow 1$ ko! $\Rightarrow$ loss $\to +\infty$

$$h_\Theta(\underline{x}) = \sigma(\underline{z}^{(L)}) \in \,]0, 1[^K$$

$$H = -\left[y \cdot \ln(h_\Theta(\underline{x})) + (1 - y) \ln(1 - h_\Theta(\underline{x})) \right]$$



Cross entropy (II) $\to$ is a measure of SURPRISE

$$J(\Theta) = -\frac{1}{m} \sum_{i=1}^{m} \sum_{k=1}^{K} y_k^{(i)} \ln\left[h_\Theta(\underline{x}^{(i)})_k\right] + (1-y_k^{(i)}) \ln\left[1-h_\Theta(\underline{x}^{(i)})_k\right]$$

$$+ \underbrace{\frac{\lambda}{2m} \sum_{j \neq 0} \left(\Theta_{ij}^{(\ell)}\right)^2}_{\text{weight decay / regularisation}}$$

$$h_\Theta(\underline{x}) = \sigma(\underline{z}^{(L)}) \in \left]0,1\right[^K$$

Softmax + negative log-likelihood $= $ CROSS-ENTROPY ($\sigma(\cdot)$ +)

$$0 < \underbrace{\text{softmax}(\underline{z}^{(L)})_k}_{\in\,]0,1[^K} = \frac{\exp(z_k^{(L)})}{\sum_{k=1}^{K} \exp(z_k^{(L)})} < 1$$

$$\sum_{k=1}^{K} \text{softmax}(\underline{z}^{(L)})_k \equiv 1$$

logSoftMax is used for numerical robustness

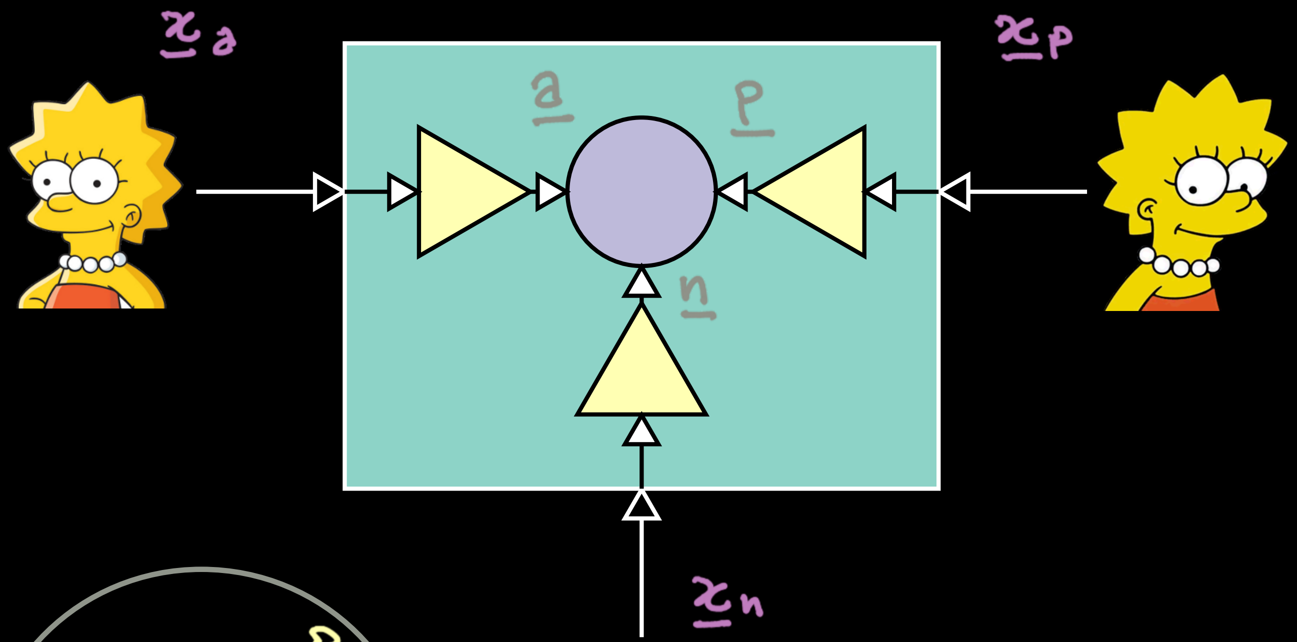
$$J(\Theta) = -\frac{1}{m}\sum_{i=1}^{m} \ln\left[h_\Theta(\underline{x}^{(i)})_{y^{(i)}}\right] =$$

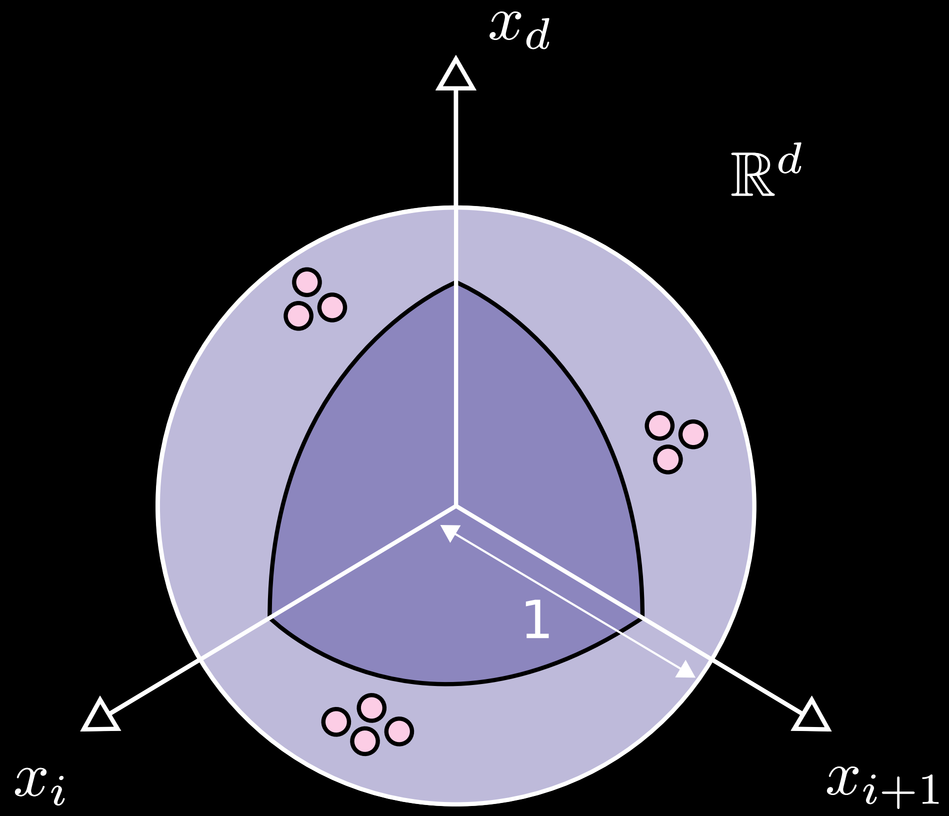
$$= \frac{1}{m}\sum_{i=1}^{m} \ln\left(\sum_{k=1}^{K}\exp(z_k^{(L)})\right)^{(i)} - \left(z_y^{(L)}\right)^{(i)} \quad \left(+\frac{\lambda}{2m}\sum_{j\neq 0}\left(\Theta_{ij}^{(\ell)}\right)^2\right)$$

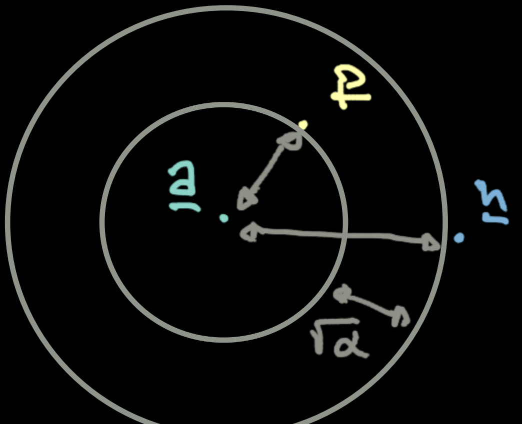



Triplet embedding loss

$\underline{a} := h_\theta(\underline{x}_a)$, same for $\underline{p}$ and $\underline{n}$







$$J(\Theta) = \frac{1}{m} \sum_{i=1}^{m} \left[\underbrace{\| \underline{a} - \underline{p} \|^2 + \alpha} - \underbrace{\| \underline{a} - \underline{n} \|^2} \right]^+$$