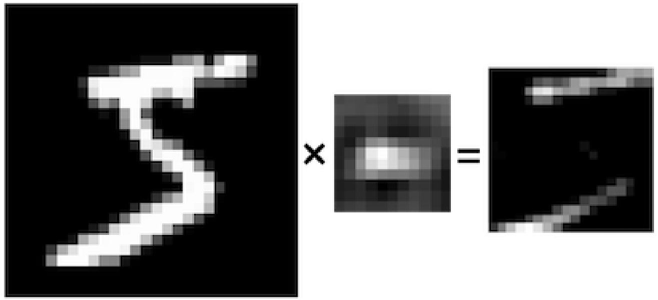
Introduction to Deep Learning

CNNs Part I



André Ofner, M.Sc.

ofner@uni-potsdam.de

29. October 2018

Schedule

- Recap (Kristin, Tunc)
- Loss functions (cross-entropy, MSE)
- Programming MLP & visualization
- CNN Quiz

Universität Potsdam

Last time on IDL

MLPs, gradient descent, backprop

Loss functions: Cross-entropy and MSE

Loss functions

- measure the quality of network output values
- With respect to the difference between
    - output unit activations <-> ground truth labels (targets)

labels (ground truth

input

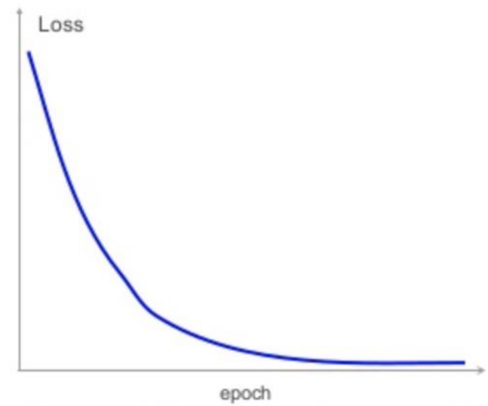
$$\mathcal{L}(w) = distance(f_\theta(x), y)$$

error

parameters (weights, biases)

Loss functions

- loss functions are used to guide the learnable parameters towards the desired performance
- by minimization of the loss function's value
- *average* loss is computed as the average of of each example (or *batch* loss)



labels (ground truth

input

$$\mathcal{L}(w) = distance(f_\theta(x), y)$$

error

parameters (weights, biases)

Types of loss functions

- Depending network and task type:
- Classification
  - categorical output
  - e.g. *cross-entropy loss*

$$-\sum_{c=1}^{M} y_{o,c} \log(p_{o,c})$$

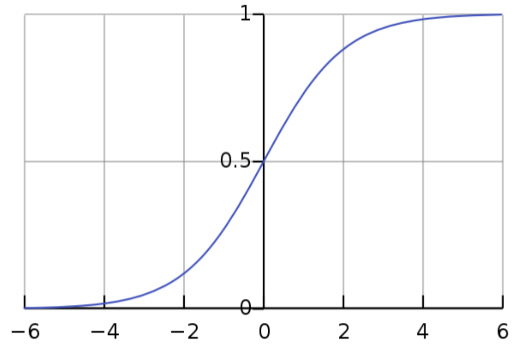
- Regression
  - continuous, numeric output
  - e.g. *mean absolute error* (MAE, L1)
  - e.g. *mean squared error* (MSE, L2)

$$MAE = \frac{\sum_{i=1}^{n} | y_i - \hat{y}_i |}{n}$$

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}$$

Output activation functions

- transform output unit vector for loss computation (in training)
- transformed outputs are also used for prediction (after training)



- **Sigmoid** function
  - transforms a vector into range (0, 1)
  - applied independently to each element
- **Softmax** function
  - sum of categorical probabilities is 1

  → cross-entropy loss

$$f_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}}$$

Mean Squared Error (MSE, L2)

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}$$

- Sum of squared errors between predictions and ground truth
- squared error is always positive
- basic loss function for regression tasks (continous, numeric variables)

Cross-entropy

$$-(y\log(p) + (1-y)\log(1-p))$$

- categorical output: probability value between 0 and 1
- multiplies the log of the predicted probability with the ground truth

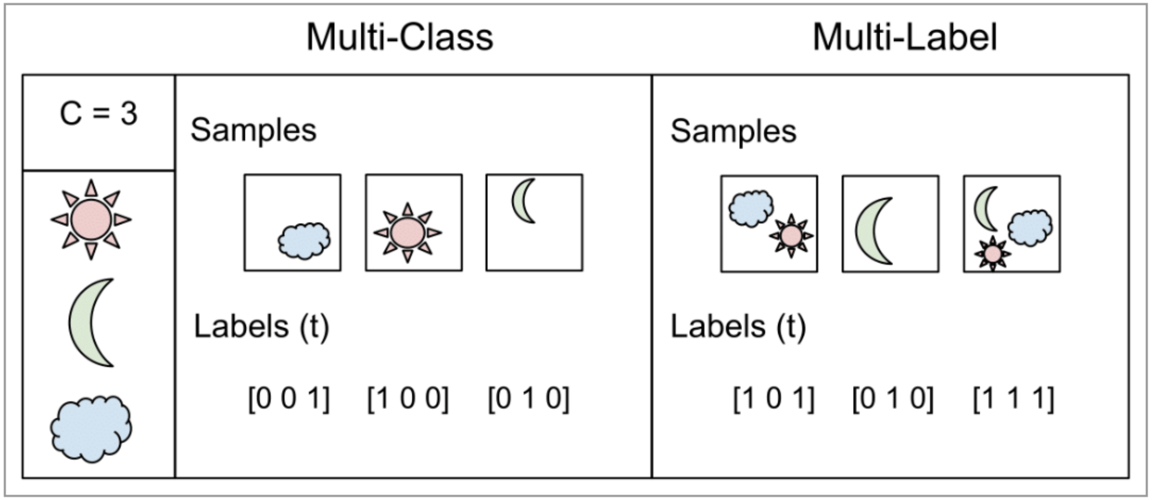
- when the ground truth is 0 (y(i) = 0), then the first part is zero
- when ground truth is 1 (y(i) = 1), second part is zero

→ penalizes strongly the predictions that are confident and wrong

Cross-entropy

$$-\sum_{c=1}^{M} y_{o,c} \log(p_{o,c})$$

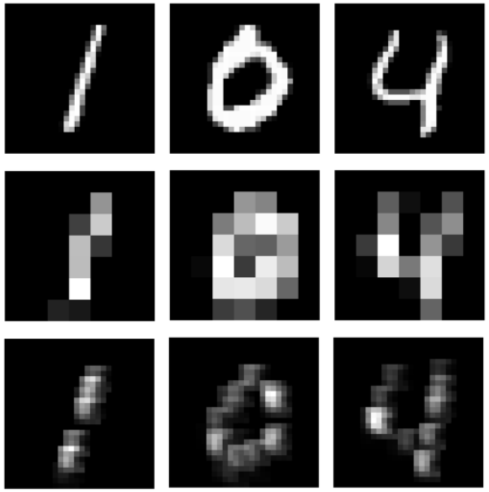
- Multiple classes → sum of losses for each class label



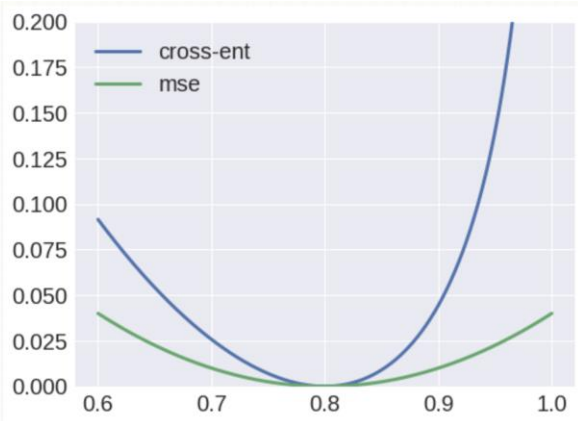
gombru.github.io/2018/05/23/cross_entropy_loss/

Cross-entropy or MSE?

Example: comparing image data



- applicable loss functions:
  - cross entropy (categorical, binary)
  - mean squared error



- important factors:
  - assumptions about data distribution
  - task / goal
  - activation function

Programming MLP & visualization

Student demo & open discussion

Convolution Quiz

in 5 groups

Assignments until next week

- Reading:
  CNNs Part II
- Programming exercise:
  CNN for MNIST
  Filter visualization

Slides & assignments on:
https://mlcogup.github.io/idl_ws18/