Deep Learning

Exercise 4: Multi-Output Networks and Batch Processing

Room: **BIN-1-B.01**
Instructor: Manuel Günther
Email: guenther@ifi.uzh.ch
Office: AND 2.54

Friday, March 18 2022

Outline

Outline

Dataset

Dataset Contents

- Student achievement in secondary education of Portuguese schools
- Different data available: demographics, school-related, social
- Task: predict grades of three grading periods
  → Multi-target regression with 3 outputs in range $[0, 20]$
- Two different subjects are evaluated: "mat" and "por"
  → 395 samples for "mat" and 649 samples for "por"

Dataset URL

https://archive.ics.uci.edu/ml/datasets/Student+Performance

Dataset

Task 1: Data Set Loading

- Download dataset from UCI
- Extract from zip file in Python
- Read CSV file `student-mat.csv` or `student-por.csv`
- Convert binary values into $\{-1., 1.\}$
- Convert integral values to `float`
- Ignore categorical values (will see later how to use these)
- Compute Input $\mathbf{X}$: columns 1–8,13–30
- Compute Targets $\mathbf{T}$: columns 31–33

Dataset

Test 1: Data Checkup

- Is the data in the desired form and shape?
- Are the target values in range?

Task 2: Data Standardization

- Compute mean μ_d and STD σ_d
  - $\rightarrow$ For each input dimension d over whole dataset
  - $\rightarrow$ Make sure that dimensions are correct!
- Standardize input data as $x_d = \frac{x_d - \mu_d}{\sigma_d}$

Dataset

Task 3: Batch Processing

- Implement a function to turn the dataset into batches of size B
- Wrap-around: start at the beginning when reaching the end
- Maybe shuffle data in between epochs
- Implement as a generator function (using `yield`)

Test 2: Test Batches

- Design input and target data to test batch processing
- Extract 20 batches
  - $\rightarrow$ Test if the size and content is correct
  - $\rightarrow$ Make sure that the targets belong to the inputs

Gradient Descent

Task 4: Multi-Target Network

- Implement network output `network(X, Theta):` for input batch $\mathbf{X}$
- Compute activation $\mathbf{A}$ as in the lecture
- Hidden unit output $\mathbf{H}$ using $\tanh$ activation function
  - $\rightarrow$ Assure that the hidden bias neuron is set to 1
- Network output $\mathbf{Y}$
- Function returns both $\mathbf{Y}$ and $\mathbf{H}$

Task 5: Loss Function

- We use $\mathcal{J}^{L_2} = \frac{1}{B}\|\mathbf{Y} - \mathbf{T}\|_F^2$ as loss
- Function takes network output $\mathbf{Y}$ and targets $\mathbf{T}$ as parameters

Gradient Descent

Task 6: Gradient Implementation

- Compute gradient as `gradient(X, T, Y, H, Theta):`
- Do not call network, use provided $\mathbf{Y}$ and $\mathbf{H}$
- Compute gradients: $\nabla_{\mathbf{W}^{(1)}}, \nabla_{\mathbf{W}^{(2)}}$ according to lecture
  - $\rightarrow$ Make sure that they are normalized correctly
  - $\rightarrow$ Assure that you use the derivative of $\tanh$!

Gradient Descent

Task 7: Gradient Descent

- Run for 10'000 epochs
  - $\rightarrow$ How many batches does this translate to?
- For each batch $(\mathbf{X}, \mathbf{T})$:
  1. Compute network output for current batch
  2. Compute and store loss value in list
  3. Compute the gradient
  4. Perform weight update:

$$\mathbf{W}^{(1)} -= \eta \nabla_{\mathbf{W}^{(1)}}$$
$$\mathbf{W}^{(2)} -= \eta \nabla_{\mathbf{W}^{(2)}}$$

- Return list of loss values

Gradient Descent

Run Gradient Descent Twice

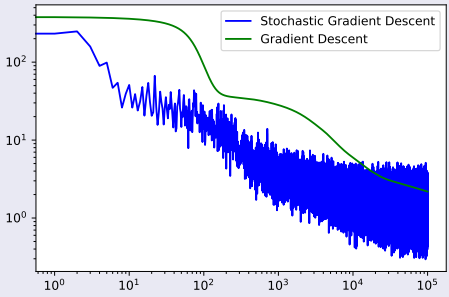
- Select an appropriate number of hidden neurons K
- Initialize parameters using Xavier method: $w \in \left[-\frac{1}{\sqrt{D}}, \frac{1}{\sqrt{D}}\right]$
- Run stochastic gradient descent with $B = 16$ and $\eta = 0.001$
- Run gradient descent with the same $\eta = 0.001$
  - $\rightarrow$ How to achieve this without changing the implementation of the function from Task 7?
- Make sure to use the same initial weights for both runs
- Store loss values from both processes

Evaluation

Task 9: Loss Progression Plot

- Plot loss of SGD as line plot
  - → Plot loss of one batch per epoch
- Plot loss of GD as line plot

Exemplary Loss Progression



Evaluation

Task 10: Example Evaluation

- Imagine one example student
  → Provide all parameters
- Predict three grades using SGD-optimized parameters

Task 11: Influence of Data Dimensions

- Take your example student and modify some details
- Check how this influences network output:
  1. male, female (index 2)
  2. paid classes (index 14)
  3. romantic relationship (index 19)
  4. daily alcohol (index 23)
- Is there a difference between courses: `"mat"` or `"por"`