Computer Science Department, University of Bonn

Knowledge Graph Analysis

Exercise Sheet 3

Dr. Hamed Shariat Yazdi, Prof. Jens Lehmann

November 15, 2018

1 In Class

1. **Classification and Regression**
   a) Give an example for a classification and a regression problem one could solve with supervised machine learning.
   b) Do you know a loss function that could be used for a classification problem? What kind of loss function could be used for a regression problem?

2. **Underfitting and Overfitting**

   You are training three classifiers on a training set. The plots in Figures 1.1, 1.2 and 1.3 show the evolution of their classification error on training and test set during training. Do the models underfit, overfit or model the data well?

3. **Independence of Random Variables**
   a) Decide if the random variables in the following scenarios are independent or not.
      - You are rolling two dices. The random variables X and Y are modeling the result of the first and the second dice, respectively, and Z is the sum the results of both dices. What is the probability $P(X = 6, Y = 3)$? Are X and Y independent? What is the probability $P(X = 5, Z = 9)$? Are X and Z independent?

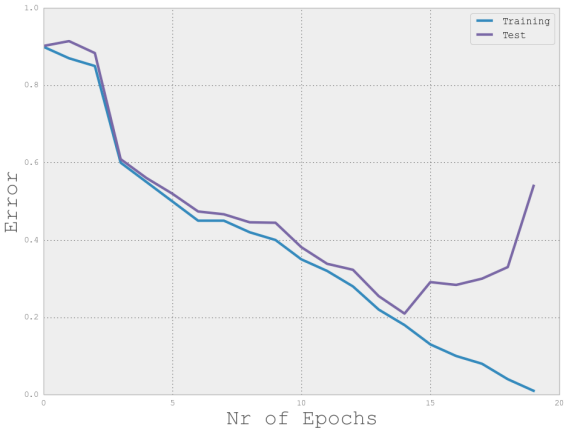

Figure 1.1: Model 1

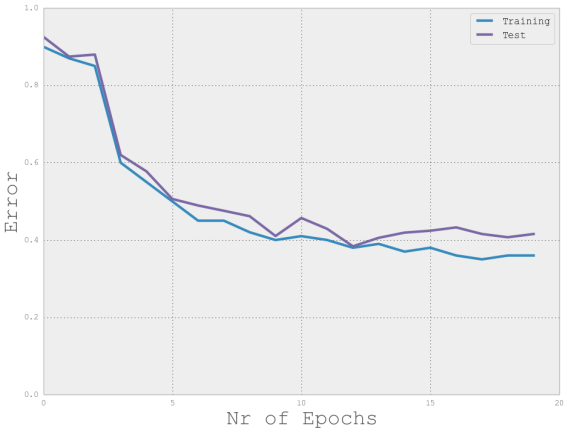

Figure 1.2: Model 2

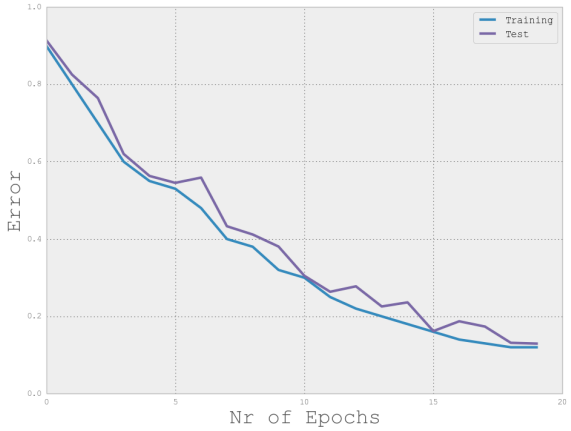

Figure 1.3: Model 3

- Two cards are drawn with replacement from a deck of 32 cards (16 black, 16 red). X and Y model the color of the fist and second card drawn, respectively. What is the probability of drawing a red card on the first trial and a red card on the second trial? Are X and Y independent? How are the answers, if the cards are drawn without replacement?

b) Let X and Y be two continuous random variables with probability densities:

$$P_X(x) = \begin{cases} \frac{x^2}{21}, & \text{if } 1 \leq x \leq 4 \\ 0, & \text{otherwise} \end{cases}$$

$$P_Y(y) = \begin{cases} y, & \text{if } 0.5 \leq y \leq 1.5 \\ 0, & \text{otherwise} \end{cases}$$

If X and Y are independent, what is their joint distribution function?

c) Let X and Y be two discrete random variables with state space $\{0, 1\}$. Let their joint probability mass function be:

$$P_{X,Y}(x, y) = \begin{cases} \frac{1}{8}, & \text{if } x = 0, y = 0 \\ \frac{3}{8}, & \text{if } x = 0, y = 1 \\ \frac{3}{8}, & \text{if } x = 1, y = 0 \\ \frac{1}{8}, & \text{if } x = 1, y = 1 \end{cases}$$

Are X and Y independent?

4. **Statistical properties of KGs and statistical relational learning (SRL) tasks**

   a) Give an example for block structure, homophily, and long-rage statistical dependencies in Knowledge Graphs.

   b) What are typical tasks in SRL?

5. **Score-based and Probabilistic models for KGA**

   a) What is the difference between score-based and probabilistic models?

   b) As discussed in the lecture, score-based models can be transferred into probabilistic models via Platt scaling[1]

$$P(y_{ijk} = 1 | f(e_i, r_j, e_k)) = sig_\epsilon(f(e_i, r_j, e_k))$$

   with

$$sig_\epsilon(x) = \begin{cases} \frac{\epsilon}{e} \exp(\frac{x}{\epsilon}), & \text{if } x \leq \epsilon \\ x, & \text{if } \epsilon \leq x \leq 1 - \epsilon \\ 1 - \frac{\epsilon}{e} \exp(\frac{1-x}{\epsilon}), & \text{if } x \geq 1 - \epsilon \end{cases}$$

   .

   - Draw a sketch of $sig_{\frac{1}{2}}$ and $sig_{\frac{1}{4}}$.
   - How can ϵ be estimated in practice?

2 At Home

1. To familiarise yourself with Python and its machine learning libraries for the next exercises, please go through the tutorial at `http://scikit-learn.org/stable/tutorial/basic/tutorial.html`.

[1] J. C. Platt, "Probabilities for SV Machines" in Advances in Large Margin Classifiers. MIT Press, pp. 61–74, 1999