Ludwig-Maximilians-Universitaet Muenchen
Institute for Informatics
Prof. Dr. Volker Tresp
Gregor Jossé
Johannes Niedermayer

03.05.2015

Machine Learning and Data Mining

Summer 2015

Exercise Sheet 7

Presentation of Solutions to the Exercise Sheet on the 10.06.2015

Exercise 7-1 Multiple Choice

Decide whether the following statements are true or false.

(a) When rolling a fair die with numbers $\in \{1, ..., 6\}^2$ the events "the first result is even" and "the second result is odd" are independent. Yes ☒ No □

(b) The expected value of a random variable is a non-linear function. Yes □ No ☒

(c) The bias of an estimator is the squared difference between the true parameter and the expected value of the parameter estimate (averaged over many datasets). Yes □ No ☒

(d) Regularization reduces the training error of a least squares model. Yes □ No ☒

(e) $k(x, y) = k_1(x,y) \cdot k_2(x,y)$ is a legitimate kernel function for two legitimate kernel functions k_1, k_2. Yes ☒ No □

(f) The product $\text{tf}_w \cdot \text{idf}_w$ is always greater than 0. Yes □ No ☒

(g) The marginal distribution of a random variable X can be computed from a joint distribution of X and another random variable Y. Yes ☒ No □

(h) A neural network can be used to linearly separate classes. Yes ☒ No □

(i) It is possible to compute the solution to a linear regression problem in closed form. Yes ☒ No □

(j) There exist homogeneous polynomial kernels with infinite-dimensional feature maps. Yes □ No ☒

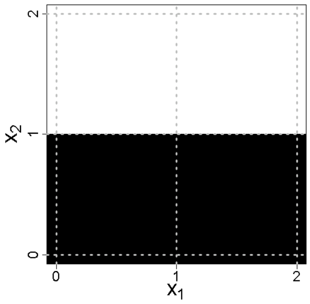
Exercise 7-2 (8 Punkte)

The Perceptron

Let $\mathbf{x} \in \mathbb{R}^M$ be a data point. Let the output of the classification of a perceptron be defined as $\hat{y} = \text{sign}(h_{\mathbf{x}}) = \text{sign}\left(\sum_{j=0}^{M} w_j\, x_j\right)$, where $x_0 = 1$ is constant.

In the following figures, the black regions are to be mapped onto $\hat{y} = 1$, the white regions are to be mapped onto $\hat{y} = -1$. Which of the mappings are possible, considering the above defined perceptron of input dimension $M = 2$?.

(a)

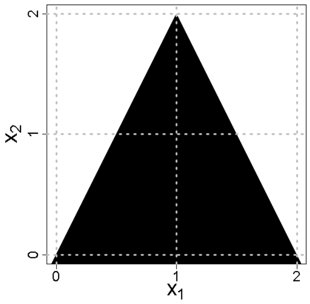


Ist moglich, da linear separierbar: Trennebene ist $x_2 = 1$, also $w_1 = 0 \Rightarrow w_2 = -\frac{w_0}{w_2}$

Lege fest $w_1 = 1 \Rightarrow \mathbf{w} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$.

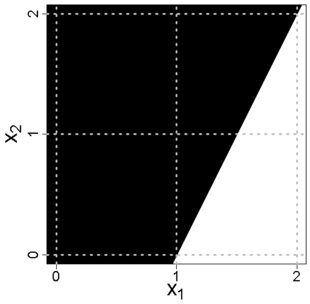
Der Test $\hat{y} = \text{sign}(h_{\mathbf{x}})$ (z.B. mit $\mathbf{x} = \binom{0}{0}, y = 1$) ergibt eine stimmige Hyperebene. Sonst musste man noch die Vorzeichen verdrehen.

(b)



Ist nicht moglich, da eine Hyperebene nicht ausreicht, um den Verlauf der Klassengrenze zu beschreiben. Es bedurfte einer zusatzlich eingefugten Fallunterscheidung um das Vorzeichen von w_1 bei 1 umzukehren.

(c)



Ist moglich, da linear separierbar: $y = ax + t = 2x - 2 \Leftrightarrow x_2 = 2x_1 - 2 = -\frac{w_1}{w_2}x_1 - \frac{w_0}{w_2}$.

Lege fest $w_0 = 1$. $\frac{w_0}{w_2} = 2 \Rightarrow w_2 = \frac{1}{2}$, und $-\frac{w_1}{w_2} = 2 \Rightarrow w_1 = -.5 \cdot 2 = -1$.

$\Rightarrow \mathbf{w} = \begin{pmatrix} 1 \\ -1 \\ 0.5 \end{pmatrix}$. Test $\hat{y} = \text{sign}(h_{\mathbf{x}})$ (z.B. mit $\mathbf{x} = \binom{2}{0}, y = 1$) passt wieder.

Exercise 7-3 Convolutional Neural Networks

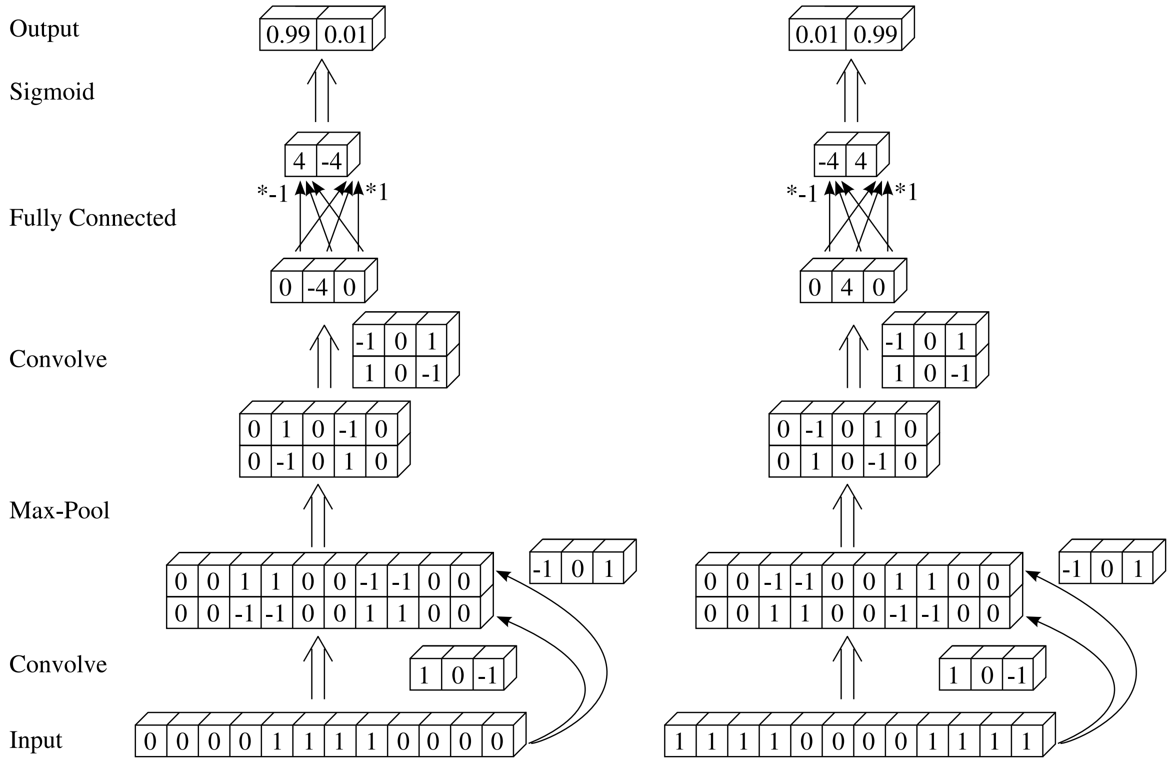
In this exercise we address a convolutional neural network (CNN) with one-dimensional input. While two-dimensional CNNs can be used for example for grayscale images, one-dimensional CNNs could be used for time-series such as temperature or humidity readings. Concepts for the 1D-case are equivalent to 2D networks. We interpret data in our network as three-dimensional arrays where a row denotes a feature map, a column denotes a single dimension of the observation, and the depth of the array represents different observations. As we will only work with a single input vector, the depth will always be one.

Let the following CNN be given:

- Input I: Matrix of size $1 \times 12 \times 1$. We therefore have an input with twelve dimensions consisting of a single feature map.
- First convolutional layer with filters $F_0^1 = (-1{,}0{,}1)$ and $F_1^1 = (1{,}0, -1)$ that generates two output feature maps from a single input feature map. Use *valid* mode for convolutions.
- Max-pooling layer with stride 2 and filter size 2. Note that max-pooling pools each feature map separately.
- Convolutional layer with convolutional kernel $F_0^2 = ((-1{,}0{,}1),(1{,}0, -1))$ of size $2 \times 3 \times 1$.
- Fully connected layer that maps all inputs to two outputs. The first output is calculated as the negative sum of all its inputs, and the second layer is calculated as the positive sum of all its inputs.
- Sigmoidal activation function

Calculate the response of the CNN for the two inputs $(0{,}0{,}0{,}0{,}1{,}1{,}1{,}1{,}0{,}0{,}0{,}0)$ and $(0{,}0{,}0{,}0{,}1{,}1{,}1{,}1{,}0{,}0{,}0{,}0)$.

Possible Solution:


Output
0.99 | 0.01
0.01 | 0.99
Sigmoid
4 | -4
-4 | 4
*-1
*1
Fully Connected
0 | -4 | 0
0 | 4 | 0
Convolve
-1 | 0 | 1
1 | 0 | -1
0 | 1 | 0 | -1 | 0
0 | -1 | 0 | 1 | 0
0 | -1 | 0 | 1 | 0
0 | 1 | 0 | -1 | 0
Max-Pool
0 | 0 | 1 | 1 | 0 | 0 | -1 | -1 | 0 | 0
0 | 0 | -1 | -1 | 0 | 0 | 1 | 1 | 0 | 0
0 | 0 | -1 | -1 | 0 | 0 | 1 | 1 | 0 | 0
0 | 0 | 1 | 1 | 0 | 0 | -1 | -1 | 0 | 0
-1 | 0 | 1
1 | 0 | -1
Convolve
Input
0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0
1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1