Deep Learning for Speech and Language Processing
Exercise Sheet 2: Basic Machine Learning

Maximilian Schmidt, Pascal Tilli, Dirk Väth, Ngoc Thang Vu

November 9th, 2020

Deadline

Please submit your solutions until Monday, November 16th, 2020, noon. You may also see the deadline at any time on the website under *Exercises > Upload*.

Submission Format

Calculation Tasks

Upload (only solutions!) to https://dlcourse.ims.uni-stuttgart.de/ using your account.

Theory Tasks

You don't have to submit this part, but it will help you with the contents of this course.

Linear Regression

Exercise 1.
Theory Task For the linear regression, the output for sample j is predicted as

$$\hat{y}^j = (\sum_{i=1}^{n} \tilde{w}_i \tilde{x}_i^j) + b = \begin{bmatrix} \tilde{x}_1^j & \tilde{x}_2^j & \cdots & \tilde{x}_n^j \end{bmatrix} \begin{bmatrix} \tilde{w}_1 \\ \tilde{w}_2 \\ \vdots \\ \tilde{w}_n \end{bmatrix} + b = \begin{bmatrix} \tilde{x}_1^j & \tilde{x}_2^j & \cdots & \tilde{x}_n^j & 1 \end{bmatrix} \begin{bmatrix} \tilde{w}_1 \\ \tilde{w}_2 \\ \vdots \\ \tilde{w}_n \\ b \end{bmatrix} = {x^j}^T w \tag{1}$$

where n denotes the number of features and w are the weights.

For m samples we get

$$\begin{bmatrix} \hat{y}^1 \\ \hat{y}^2 \\ \vdots \\ \hat{y}^m \end{bmatrix} = \hat{y} = \begin{bmatrix} x_1^1 & x_2^1 & \cdots & x_n^1 & 1 \\ x_1^2 & x_2^2 & \cdots & x_n^2 & 1 \\ \vdots & \vdots & \ddots & \vdots & 1 \\ x_1^m & x_2^m & \cdots & x_n^m & 1 \end{bmatrix} \begin{bmatrix} \tilde{w}_1 \\ \tilde{w}_2 \\ \vdots \\ \tilde{w}_n \\ b \end{bmatrix} = Xw. \tag{2}$$

To compute optimal weights we have to minimize our loss function with true labels y:

$$w^* = \arg\min_w \mathrm{MSE}(\hat{y}, y) = \arg\min_w \frac{1}{m} \sum_{i=1}^{m} (\hat{y} - y)_i^2 \tag{3}$$

Since this is a quadratic function, the analytical solution is given by setting the derivative wrt w equal to 0 and solve for w:

$$\frac{\mathrm{d}}{\mathrm{d}w} \frac{1}{m} \sum_{i=1}^{m} (\hat{y} - y)_i^2 \stackrel{!}{=} 0 \Rightarrow w = ? \tag{4}$$

Compute the optimal weights $w*$.

Exercise 2.
Calculation Task

(1) Compute the optimal parameters

$$w = (X^T X)^{-1} X^T y \tag{5}$$

of the linear regression model for the given data, where $X \in \mathbb{R}^{m \times n}$ corresponds to the input data as a matrix with m samples and n features and $y \in \mathbb{R}^n$ corresponds to the label vector (the house prices in this case).[1]

Sample id	Features			Price (in 100K $)
	Overall Quality	Overall Condition	# Bedrooms	
0	5	9	3	138
1	3	4	4	147
2	5	6	1	222
3	6	8	3	282

The analytical solution for the inverse of a 3×3 matrix is given by

$$M^{-1} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}^{-1} = \frac{1}{\det(M)} \begin{bmatrix} (ei - fh) & -(bi - ch) & (bf - ce) \\ -(di - fg) & (ai - cg) & -(af - cd) \\ (dh - eg) & -(ah - bg) & (ae - bd) \end{bmatrix} \tag{6}$$

The determinant of M can be computed by applying the rule of Sarrus:

$$\det(M) = a * (ei - fh) + b * -(di - fg) + c * (dh - eg) \tag{7}$$

<u>Omit the bias in this exercise!</u>

(2) Compute the predictions $y \in R^2$ using the previous obtained, optimal parameters w for following houses.

Overall Quality	Overall Condition	# Bedrooms
4	3	2
8	10	2

[1]Note that this example omits some features hence the given samples might not make much sense.