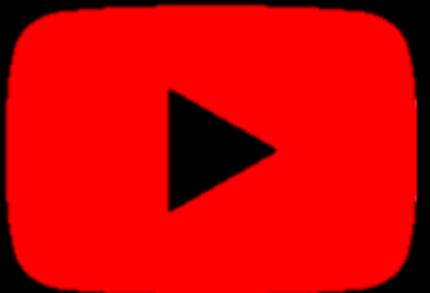


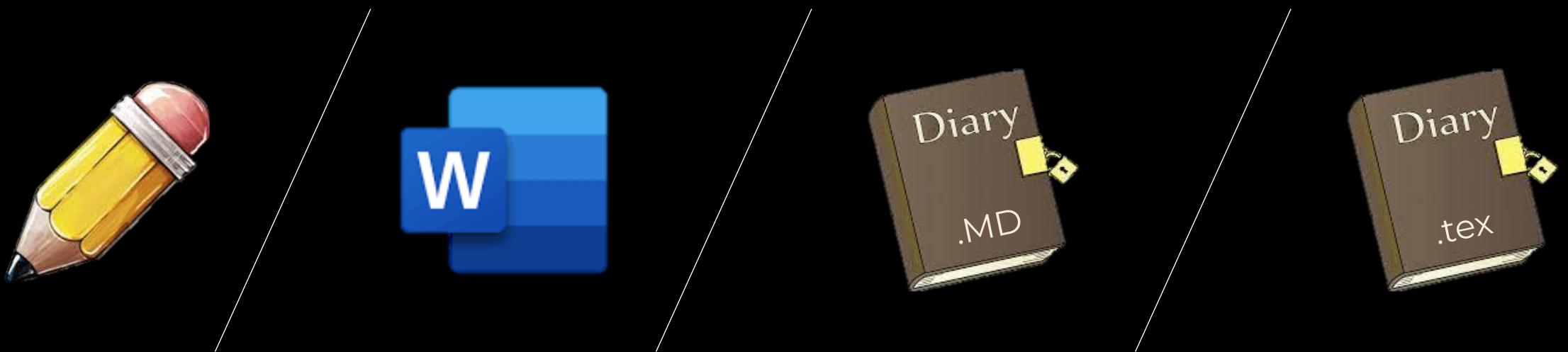
Informatik Q3 Abels



Konzepte und Anwendungen der Theoretischen Informatik



Theoretische Informatik



Machine Learning Zusammenfassung

overleaf.com/project/5d318606d6e59b73b72fa1d9

Machine Learning Zusammenfassung

Code Editor Visual Editor Normal text B I Ω % \mathbb{R} Recompile

Kernel_2.png Kernel_3.png kovex_konkav.png LearningProblemsOverview.PNG logRegPlot.png MachineLearningBesprechungBl... main.tex margin.png multReg.png RBF_Network_Architecture.png Sheet4Exercisee.PNG Sheet5Exercise2.PNG Sheet5Exercise3.PNG

File outline

- Combined:
- Exercises:
 - Sheet 4 Exercise e
 - Sheet 5 Exercise 3
 - Sheet 5 Exercise 2
- Regression
 - Linear Regression
 - 1-Parameter Linear Regress...
 - Maximum-Likelihood Estima...
 - Multivariate Regression
 - Corresponding Assignments
 - Ridge Regression
 - Logistic Regression
 - Generative vs. Discriminative M...
 - Determining the weights by ...
 - Further Remarks on Logistic ...
 - Corresponding Assignments

using a given set of training instances $\{x_{\text{training}} \subset \mathbb{R}^{m+1}\}$.
 421 $y = w_0 + w_1 x_1 + \dots + w_m x_m$
 422 So we assume, that the expected value of the output y , given an input x , $E[y|x]$ is linear.
 423 \hat{y}
 424 **1-Parameter Linear Regression**
 425 We assume the training data to be formed by a function

$$y_i = w x_i + noise_i$$

 426 The noise terms of the different training instances are assumed to be independent of each other and normally distributed ($\mu = 0$; σ^2 unknown). That means, $p(y|w, x)$ is normally distributed with mean wx and the same variance σ^2 . Now, we can estimate w by maximizing this probability over all training instances (x, y) ($p(w|x_1, \dots, x_n, y_1, \dots, y_n)$).
 427 \hat{y}
 428 **Maximum-Likelihood Estimation**
 429 We pose the question: "For which value of w is the data most likely to have happened?"
 430 \Leftrightarrow For what w is $p(y_1, \dots, y_n | x_1, \dots, x_n, w)$ maximized?
 431 \Leftrightarrow For what w is the sum-of-square residuals $E = \sum_{i=1}^n (y_i - w x_i)^2$ minimized?
 432 \Leftrightarrow For what w is the sum-of-square residuals $E = \sum_{i=1}^n (y_i - w x_i)^2$ minimized?
 433 The answer is $w = \frac{\sum (x_i y_i)}{\sum x_i^2}$
 434 So we simply use $y = wx$ for predicting the value of the target function for a given parameter x .
 435 \hat{y}
 436 **Multivariate Regression**
 437 Generalization of 1-Parameter Linear Regression: Instead of a single real-valued attribute, each of the R training instances consists of a real-valued, m -dimensional vector.
 438 $\begin{aligned} & \text{Generalization of 1-Parameter Linear Regression: Instead of a single real-valued attribute, each of the } \\ & \text{the } R \text{ training instances consists of a real-valued, } m \text{-dimensional vector.} \end{aligned}$
 439 $\begin{aligned} & \text{We describe the set of training instances as a } R \times m \text{ matrix, combined with a } R \text{-dimensional vector representing the set of values of the class attribute (the attribute, we want to estimate).} \\ & \text{Now, we want to determine a vector } w \text{ of weights instead of a single weight:} \end{aligned}$
 440
$$y = w x = w_1 x[1] + w_2 x[2] + \dots + w_m x[m]$$

 441 (where $x[i]$ denotes the i th column of x , which can be thought of as an attribute)
 442 Again, the probability of the observed y -values given the known values of x shall be maximized. In this case, we use
 443
$$w = (X^T X)^{-1} (X^T Y)$$

 444 as max. likelihood estimation.
 445 \newline It is possible to model a target function with a constant term by just adding another, 0th attribute, that always takes the value 1. We can then determine (learn) the corresponding weight, w_0 , as the value of the target function if we set all attributes to 0.
 446 $\begin{aligned} & \text{We describe the set of training instances as a } R \times m \text{ matrix, combined with a } R \text{-dimensional vector representing the set of values of the class attribute (the attribute, we want to estimate).} \\ & \text{Now, we want to determine a vector } w \text{ of weights instead of a single weight:} \end{aligned}$
 447
$$y = wx = w_1 x[1] + w_2 x[2] + \dots + w_m x[m]$$

 448 (where $x[i]$ denotes the i th column of x , which can be thought of as an attribute)
 449 Again, the probability of the observed y -values given the known values of x shall be maximized. In this case,
 450 $\begin{aligned} & \text{We describe the set of training instances as a } R \times m \text{ matrix, combined with a } R \text{-dimensional vector representing the set of values of the class attribute (the attribute, we want to estimate).} \\ & \text{Now, we want to determine a vector } w \text{ of weights instead of a single weight:} \end{aligned}$

With attributes x_1, \dots, x_m , predicted value y and weights w_0, \dots, w_m calculated using a given set of training instances $X_{\text{training}} \subset \mathbb{R}^{m+1}$.
 So we assume, that the expected value of the output y , given an input x , $E[y|x]$ is linear.

5.1.1 1-Parameter Linear Regression

We assume the training data to be formed by a function

$$y_i = w x_i + noise_i$$

The noise terms of the different training instances are assumed to be independent of each other and normally distributed ($\mu = 0$; σ^2 unknown). That means, $p(y|w, x)$ is normally distributed with mean wx and the same variance σ^2 . Now, we can estimate w by maximizing this probability over all training instances (x, y) .

2 Maximum-Likelihood Estimation

We pose the question: "For which value of w is the data most likely to have happened?"
 \Leftrightarrow For what w is $p(y_1, \dots, y_n | x_1, \dots, x_n, w)$ maximized?
 \Leftrightarrow For what w is $\prod_{i=1}^n p(y_i | w, x_i)$ maximized?
 \Leftrightarrow For what w is the sum-of-square residuals $E = \sum_{i=1}^n (y_i - w x_i)^2 = \sum_{i=1}^n y_i^2 - (2 \sum x_i y_i)w + (\sum x_i^2)w^2$ minimized?
 The answer is $w = \frac{\sum x_i y_i}{\sum x_i^2}$
 So we simply use $y = wx$ for predicting the value of the target function for a given parameter x .

5.1.3 Multivariate Regression

Generalization of 1-Parameter Linear Regression: Instead of a single real-valued attribute, each of the R training instances consists of a real-valued, m -dimensional vector.

$$\mathbf{x} = \begin{bmatrix} \dots & \mathbf{x}_1 & \dots \\ \dots & \mathbf{x}_2 & \dots \\ \vdots & & \vdots \\ \dots & \mathbf{x}_R & \dots \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & & \vdots & \\ x_{R1} & x_{R2} & \dots & x_{Rm} \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_R \end{bmatrix}$$

We describe the set of training instances as a $R \times m$ matrix, combined with a R -dimensional vector representing the set of values of the class attribute (the attribute, we want to estimate).
 Now, we want to determine a vector w of weights instead of a single weight:

$$y = wx = w_1 x[1] + w_2 x[2] + \dots + w_m x[m]$$

(where $x[i]$ denotes the i th column of x , which can be thought of as an attribute)
 Again, the probability of the observed y -values given the known values of x shall be maximized. In this case,



Tagebucheintrag





Wochenübung

- Schaffe dein persönliches Setup zur Theoretischen Informatik.
- Schaue dir meine Empfehlungen an:
 - Overleaf:
<https://www.overleaf.com/>
 - Overleaf Tutorial:
<https://www.youtube.com/watch?v=Jp0IPj2-DQA&list=PLHXZ9OQGMqxcWWkx2DMnQmj5os2X5ZR73>
 - Latex Tutorial:
<https://www.youtube.com/watch?v=hRwUjJYeHjw&list=PLuyjaM3UzoOS7zcMFaROwrg83KBR1Sui>
 - Inhalte auf Youtube:
https://www.youtube.com/watch?v=5jFKI_G5QWg&list=PLG_1tsKrsKVO2ANHX68UbrNgt7gZuH37H => NICHT ALLES!
 - Inhalte spielerisch lernen:
<https://inf-schule.de/automaten-sprachen> => NICHT ALLES!