

## Einsendeaufgabe 6 – Machine Learning

### 1. Read this paper by Pedro Domingos very carefully sentence by sentence! A Few Useful Things to Know about Machine Learning

#### - What is the meaning of generalization in the paper?

Ein wichtiges Ziel im Bereich Machine Learning ist die Generalisierung über die Trainingsdaten hinaus. Durch das Ziel Generalisierung entstehen noch weitere wichtige Konsequenzen: Nicht nur die Daten sind wichtig. Es muss darüber hinaus Wissen und Annahmen zu den Daten geben, um diese generalisieren zu können.

Dadurch können neue, unvorhersehbare Daten besser verarbeitet werden.

#### -What are the many faces of overfitting?

Es gibt verschiedene Formen von overfitting:

Bias -> immer wieder das gleiche Falsche lernen

Varicance -> die Tendenz, unabhängig vom tatsächlichen Signal zufällige Dinge zu lernen

Cross validation kann helfen das Overfitting zu minimieren, z.B. durch die Entscheidung über die Größe eines Entscheidungsbaums.

Regularization terms werden außerdem genutzt, um Classifern mehr Struktur zu geben.

Statistical significance test sollten durchgeführt werden, bevor die Struktur geändert wird. Hierbei wird geprüft, ob die Verteilung der Klasse wirklich anders ist mit und ohne eine Struktur

#### - Why do humans have problems in higher dimensions?

Menschen leben in einer 3-dimensionalen Welt. Sobald die Zahl an Dimensionen zunimmt, fällt es Menschen schwer sich diese Dinge vorzustellen, weil wir keinen Vergleich zur Realität haben.

Dadurch fällt Menschen die Entwicklung von guten Classifiern, mit vielen Dimensionen, schwer.

#### - What is feature engineering?

Feature engineering ist bereichsspezifisch. Der heilige Gral im Machine Learning ist mehr und mehr Prozesse von feature engineering zu automatisieren. Es ist der Prozess des Selektierens und Sortieren von Features. Sowie das Festlegen davon welche Features wichtig sind und welche nicht.

#### - Why does more data beat clever algorithms?

Wichtig ist die Darstellung der Daten sowie die Datenmenge für das Ergebnis. Die Qualität ist dahingehend zweitrangig. Wenn man einen sehr guten Algorithmus hat aber wenige Daten und eine schlechte Darstellung ist das Ergebnis schlechter als umgekehrt.

#### - What is ensemble learning?

Beim ensemble learning werden viele verschiedene Leaners kombiniert, um ein möglichst gutes Ergebnis aufgrund der vielen Variationen zu erhalten. Durch das kombinieren können typische Fehlerquellen wie Bias, Varianz und Rauschen minimiert werden. Womit der Algorithmus bessere Ergebnisse erzielt.

- What is accuracy in data science?

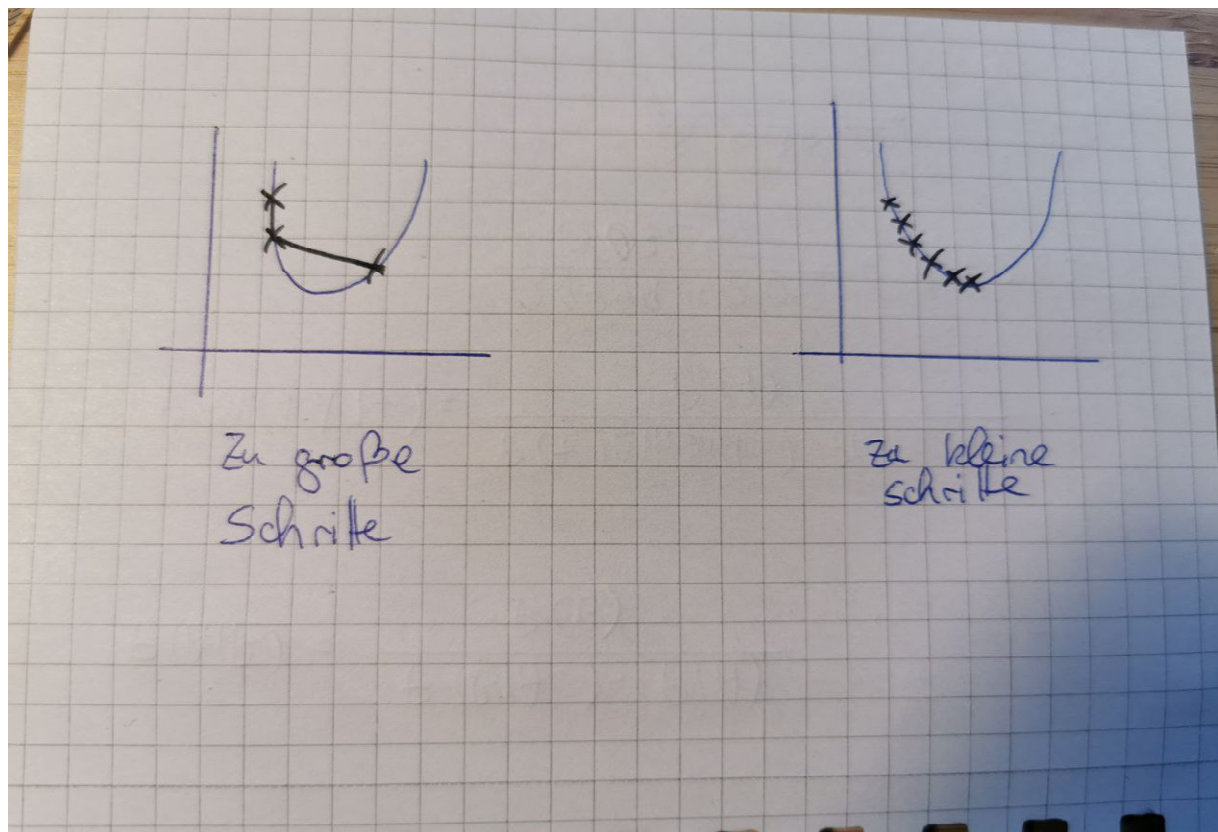
Mit der Genauigkeit soll gesagt werden, wie gut das gewünschte Ergebnis vorgesagt werden kann. Die Genauigkeit kann mit Hilfe der Berechnung der Precision sowie des Recalls beschrieben werden.

2. Follow the given resources on **gradient descent** or find new / better resources.

Describe the core idea of gradient descent in a few paragraphs. Pictures, equations or anything you need is also fine.

Der Gradient Descent ist eine numerische Methode um Geschäftsprobleme mit maschinellen Lernalgorithmen wie Regression, neuronalen Netzen, Deep Learning usw. zu lösen. Es handelt sich um einen Optimierungsalgorithmus. Es ist ein iteratives Verfahren, bei dem nach dem Punkt mit dem steilsten Abstieg gesucht wird.

Dabei dürfen die einzelnen Annäherungsschritt nicht zu groß gewählt werden, denn dabei könnte der tiefste Punkt übersprungen werden. Werden die Schritte allerdings zu klein gewählt müssen viele Berechnungen durchgeführt werden, was zu höheren Kosten führen wird. Wie Groß die einzelnen Schritte sind wird von der learning rate festgelegt.



3. Learn about **Bayesian Statistics / Algorithm** from this video [\[1\]](#) or find a better resource. Tell me what you learned including the best example you came over.

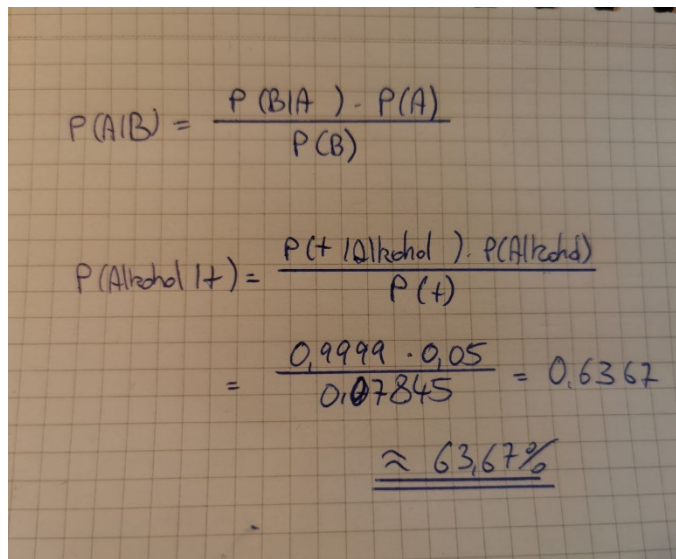
Es ist ein zentrales Werkzeug in ML und AI. Das Bayes Theorem gibt es schon seit 1770.

Mit dem Satz von Bayes lässt sich für 2 verschiedene Ereignisse A und B, die Wahrscheinlichkeit berechnen, ob ein Ereignis eintritt.

Ein Beispiel wäre ein Alkoholtest. Dieser ist in 99,9% der Fälle positiv, wenn Alkohol konsumiert wurde. Er kann aber auch positiv sein, obwohl kein Alkohol konsumiert wurde, dies geschieht in 3% er Test.

5% der getesteten Personen haben tatsächlich Alkohol getrunken. 95% haben also keinen Alkohol getrunken.

Wenn der Test nun positiv ausfällt, kann mit dem Satz von Bayes berechnet werden, wie hoch die Wahrscheinlichkeit ist, dass tatsächlich Alkohol konsumiert, wurde:



The image shows a handwritten calculation on grid paper. It starts with the general formula for Bayes' Theorem: 
$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$
 Then, it applies this to an alcohol test example. Let A be 'Alcohol consumed' and B be 'Test positive'. The calculation is: 
$$P(\text{Alkohol} | +) = \frac{P(+ | \text{Alkohol}) \cdot P(\text{Alkohol})}{P(+)}$$
 
$$= \frac{0,9999 \cdot 0,05}{0,07845} = 0,6367$$
 
$$\approx \underline{\underline{63,67\%}}$$