Technische Universität Dresden

Fakultät Maschinenwesen

Institut für Maschinenelemente und Maschinenkonstruktion

Professur für Konstruktionstechnik/CAD

**Studienarbeit**

Thema: **Untersuchung zur Vorhersage der maximalen Verschiebungen von Bauteilen auf Basis wissensbasierter Methoden**

Topic: **Evaluation of maximum displacements of components based on knowledge-based methods**

Name des Verfassers: Jie Zhao

Matrikel-Nr.: 4746248

Betreuer: Stefan Holtzhausen, Sebastian Klement

Verantwortlicher Hochschullehrer: Prof. Dr.-Ing. habil. R. Stelzer

Ausgegeben am: 01.02.2019

Einzureichen am: 01.08.2019

Inhalt

[1. Einleitung und Motivation 1](#_Toc14111226)

[2. Stand der Technik 2](#_Toc14111227)

[2.1 Maschinelles Lernen 2](#_Toc14111228)

[2.1.1 Arten und Zielstellungen 3](#_Toc14111229)

[2.1.2 Gradientenabstiegsverfahren und Verallgemeinerung 9](#_Toc14111230)

[2.2 Methoden im maschinellen Lernen 11](#_Toc14111231)

[2.2.1 Support Vector Maschine (SVM) 11](#_Toc14111232)

[2.2.2 Neuronales Netz 15](#_Toc14111233)

[3. Generierung der FEM-basierten Daten 19](#_Toc14111234)

[3.1 Vorgehensweise einer FEM-Analyse 19](#_Toc14111235)

[3.2 Automatisierung der Vorgehensweise durch SolidWorks-API mittels C# Programmierung 23](#_Toc14111236)

[4 Statistische Untersuchung 26](#_Toc14111237)

[4.1 Korrelation innerhalb des Datensatz 26](#_Toc14111238)

[4.2 Statistische Verteilung 28](#_Toc14111239)

[4.3 Verstärkung bzw. Ausgleich des Datensatz 30](#_Toc14111240)

[4.4 Auswirkung der Verstärkung des Datensatz 34](#_Toc14111241)

[5 Daten vorbereiten 36](#_Toc14111242)

[5.1 Datenbereinigung und Ausreißererkennung 36](#_Toc14111243)

[5.2 Feature Scaling 38](#_Toc14111244)

[6 Vorhersagemodell erstellen 41](#_Toc14111245)

[6.1 SVM Model für ein Balkenbauteil unter Lasten 41](#_Toc14111246)

[6.2 Neuronales Netz für W-Profil-Bauteile unter Lasten 43](#_Toc14111247)

[6.3 Stufiges Modell für W-Profil-Bauteile unter Lasten 46](#_Toc14111248)

[7 Auswertung und Bewertung 51](#_Toc14111249)

[8 Zusammenfassung und Ausblick 53](#_Toc14111250)

[9 Literaturverzeichnis 55](#_Toc14111251)

**Symbolverzeichnis**

**Formelzeichen**

|  |  |  |
| --- | --- | --- |
|  | Dimension | Beschreibung |
|  | - | Strafterm |
|  | - | Kostenfunktion |
|  | - | Anzahl der Eingangsfeature |
|  | - | Anzahl der Daten im Datensatz |
|  | - | Gewichte der Verbindungen |
|  | - | Bias |
|  | - | Realer Wert |
|  | - | Vorhersage |
|  | - | Learningrate |
|  | - | Kovarianz |
|  | - | Korrelationseffizient |
|  | - | Standardabweichung |
|  | - | Funktioneller Margin |
|  | - | Geometrischer Margin |
|  | - | Toleranz |
|  | - | 2-Norm eines Vektors |

**Abkürzungen**

|  |  |
| --- | --- |
| FEM | Finite-Elemente-Methode |
| FP | False Positiv |
| FN | False Negativ |
| MAE | Mean Absoult Error |
| MAPE | Mean Absoult Percentage Error |
| MSE | Mean Squard Error |
| NN | Neuronales Netz |
| RBF | Radial Basis Function |
| SVM | Support Vector Machine |
| TP | True Positiv |
| TN | True Negativ |

# 1. Einleitung und Motivation

Bis Mitte des 20. Jahrhunderts wurde zunächst die theoretische Grundlage der Finite-Elemente-Methode (FEM) gebildet. Dank der Fortschritte von Rechnern ist das Lösen das komplexe theoretische Gleichungssystem möglich. Heutzutage ist FEM in der festigkeitsmäßigen Auslegung von Bauteilen weit verbreitet. Die FEM basiert auf dem Lösen von unterschiedlichen Differenzialgleichungen mittels numerischen iterativen Verfahren. Ein Bauteil wird in endlich viele Elemente, die mit einfacher und nicht überlappender Form gekennzeichnet ist, aufgeteilt. Dazu spielen sowohl die Form der Elemente als auch die Größe der Elemente eine entscheidende Rolle. Deswegen erfordert FEM entsprechend qualifiziertes Fachpersonal. Das Lösen einer Variante von Bauteilen kann manchmal stundenlang dauern. Die benötigte Bearbeitungszeit ist abhängig von Geometrie und Rechenleistung. Vorteil der FEM liegt daran, dass nach einer Berechnung alle Spannungen und dazu entsprechende Verformungen herauskommen. Bei parametrisierbaren Geometriemodellen, welche individuell auf eine maximale Zielverformung hin, ausgelegt werden soll, ist die FEM aufwendig [1, 2].

Das Interesse für künstliche neuronale Netze (KNN) setzte bereits in den frühen 1940er Jahren ein, also etwa gleichzeitig mit dem Einsatz programmierbarer Computer in angewandter Mathematik [2]. Aber wegen die Beschränkung der Reichenleistung entwickelte sich die KNN sehr langsam. Bereits 1974 entwickelte Paul Werbos für seine Dissertation die Backpropagation bzw. die Fehlerrückführung [3]. Das Modell war aber erst später von einer größeren Bedeutung. Neuronales Netz (NN) ist heutzutage ein sehr erfolgreicher Ansatz aus dem Bereich des maschinellen Lernens, welcher auf Basis von verfügbaren Wissen verallgemeinert. D.h. es lernt nicht das verfügbare Wissen auswendig, sondern erkennt. Auf Grund der Fortschritte die Rechenleistung von CPU und GPU entsteht die Möglichkeit, tieferes und tieferes Neuronales Netz zu antrainieren [2]. Typische eingesetzte Beispiele von maschinellen Lernen sind Klassifikation von Bildern, Text- und Spracheübersetzung, und Empfehlungssystem. Dazu kommt spezielle Struktur des Netzes, z.B. Convolutional Neural Network (CNN) und Rekurrentes neuronales Netz (RNN).

Im Rahmen der Arbeit soll untersucht werden, ob wissensbasierte Methoden (Support Vector Maschine - SVM, Neuronale Netze - NN) ähnlich qualifizierte Aussagen treffen können, wie die numerische Simulation. Diese Arbeit konzentriert sich darauf, wie die maximale Verschiebung eines Bauteils unter bestimmten Lasten vorhergesagt werden kann. Dazu sind an mehreren parametrischen Bauteilen entsprechende SVM und NN aufzustellen, zu dimensionieren und zu validieren. Die benötigten Daten bzw. Wissen wird durch ein Addin-Modul „Simulation“ in einer CAD Software SolidWorks erzeugt. Um die Datengenerierung automatisch laufen zu können, ist die Anruf von SolidWorks durch „Application Programming Interface“ (API) in C# nötig. Die entsprechenden Wertebereiche für Bauteildimensionen und Lasten sind sinnvoll einzuschränken.

Um die vollständig Vorgehensweise zu klären, wird in der Arbeit 3 parametrischen Bauteile als Beispiele genommen, nämlich Balken, L-Profil und W-Profil (siehe Abb. 1).

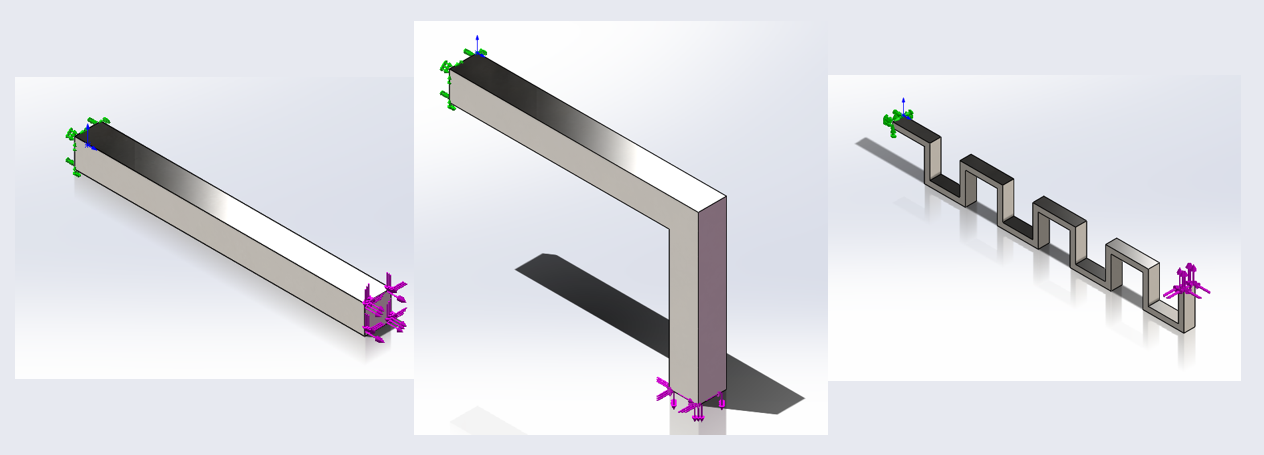


Abb.1: Beispiele der Bauteile

(links) Balken, (mittel)L-Profil, (rechts) W-Profil

Die Arbeit ist anhand der Vorgehensweise der Erzeugung eines Vorhersagemodells eingegliedert. Zuerst müssen die benötigten Daten bzw. Wissen durch ein automatisiertes Skript erfassen werden. Danach wird die ansammelte Daten in einem bestimmten Regeln erkundet. Dann folgt das Erstellen eines Vorhersagemodells. Dazu sind mehre Modelle mit unterschiedlichen Strukturen und Parametern sich voneinander verglichen. Im letzten Teil ist die Bewertung und Vertiefungsrichtung dargestellt.

# 2. Stand der Technik

## 2.1 Maschinelles Lernen

Ziel der Arbeit ist es ein wissensbasiertes Vorhersagemodell mit ausreichender Genauigkeit die maximalen Verschiebungen eines Bauteils unter bestimmten Lasten zu erstellen. Normalerweise ist sogenannte wissensbasierte Modell ein obenliegender Begriff. Darunter stehen Big Data, Künstliche Intelligenz, usw. Abb. 2 zeigt die Gliederung des Begriffs im Bereich eines wissensbasierten Modells. Darauf stehen Support Vector Machine (SVM) und Neuronales Netz (NN) im Fokus, weil die in der Arbeit zum Einsatz kommen.

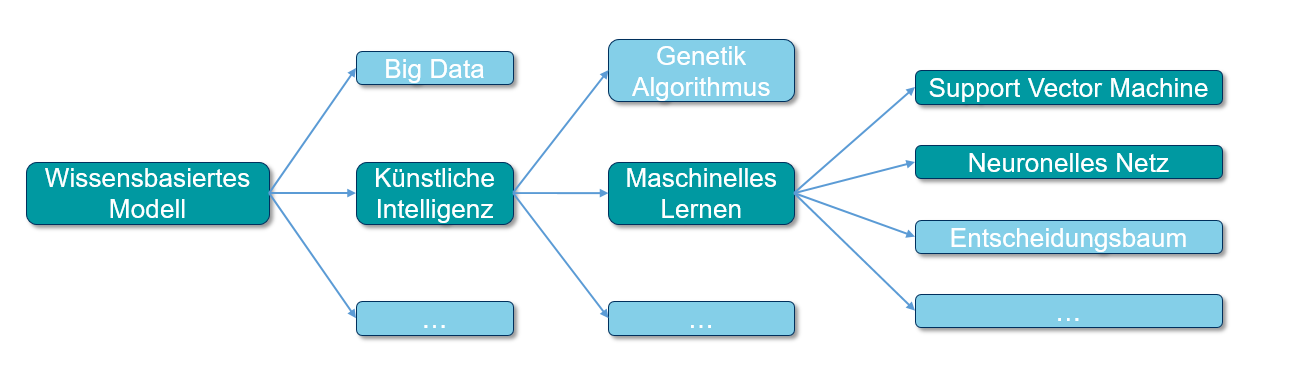


Abb. 2: Gliederung des Begriffs

Im Ingenieurbereich wirken sich die in Abb. 2 stehende Techniken als Werkzeuge, wodurch können Probleme mit einem kreativen Lösungsweg gelöst werden. Dazu verwendet man im Wesentlichen die vordefinierten Bibliotheken, welche von Informatiker und Mathematiker zusammengestellt werden. Einige typische Beispiele dazu sind TensorFlow (von Google), Sckite-Learn, PyTorch (von Facebook).

TensorFlow ist ein Framework zur datenstromorientierten Programmierung. Es wird aus Python-Programmen heraus benutzt und in Python und C++ implementiert. Populäre Anwendung findet TensorFlow im Bereich des maschinellen Lernens [4]. In der Arbeit wird TensorFlow auch angewendet, das Vorhersagemodell zu erstellen.

Unter TensorFlow steht ein Teil Keras zur Verfügung. Keras vereinfacht die Programmierung und ermöglicht einen schnellen Anfang bzw. Keras verhält sich wie modulare Bauteile, damit kann man unterschiedliche Modelle bauen. Deswegen spielt es sowohl im Ingenieurbereich als auch in der Forschung eine wesentliche Rolle [5].

Scikit-Learn ist eine freie Software-Bibliothek zum maschinellen Lernen für die Programmiersprache Python [6]. In der Arbeit wird die als Hilfsfunktion zum Trainieren und Testen des Vorhersagemodells eingesetzt.

Neben die TensorFlow und Scikit-Learn sind noch ein paar Bibliotheken benötigt, z.B. Numpy, Pandas, Seaborn usw. Die behandeln sich um sowohl Datenstruktur und Datenvisualisierung als auch Datenbearbeitung. Großer Vorteil davon ist es, dass die typischen Bibliotheken zum maschinellen Lernen dazu schon angepasst werden.

2.1.1 Arten und Zielstellungen

Im maschinellen Lernen treten 3 Type des Datensatz häufig auf, nämlich Trainingsdatensatz, Validierungsdatensatz und Testdatensatz. Die haben unterschiedliche Funktion beim Erstellen eines Vorhersagemodells. Daher sind die Drei nicht miteinander umtauschbar.

Trainingsdatensatz: Der besitzt große Anzahl von Daten, die zum Trainieren des Modells dienen.

Validierungsdatensatz: In einer Gittersuche wird eine Beurteilung eines Modells auf diesem Datensatz benötigt, womit erhalt man die Reihenfolge der Modelle in der Gittersuche.

Testdatensatz: Zur Bewertung eines Modells zum Schluss auf diesem Datensatz durchzuführen. Der darf niemals vorher ins Modell zum Trainieren u./o. Validieren verwendet werden.

Alle Drei lassen sich von einer gemeinsamen Quelle erfassen. In der Arbeit sind sie durch einer automatisierten FEM-Analyse angesammelt (siehe Kap. 3.2). Die alle Modellgüte-Darstellungen in der Arbeit beruhen auf Testdaten.

Begriff von maschinellen Lernen ist die künstliche Generierung von Wissen aus Erfahrung. Ein künstliches System lernt aus Beispielen (gefütterte Wissen) und kann die nach Beendigung der Lernphase verallgemeinern.

Alpaydin (2016) sagte so, maschinelles Lernen ist ein Teilgebiet der künstlichen Intelligenz. Durch das Erkennen von Mustern und Gesetzmäßigkeiten generieren Computersysteme selbständig Wissen aus vorliegenden Datenbeständen [7].

Zusammenfassend kann man sagen, dass maschinelles Lernen auf Herausfinden einer Funktion abzielen, die die Zusammenhänge zwischen Aus- und Eingangsdaten beschreiben (siehe Abb. 3).

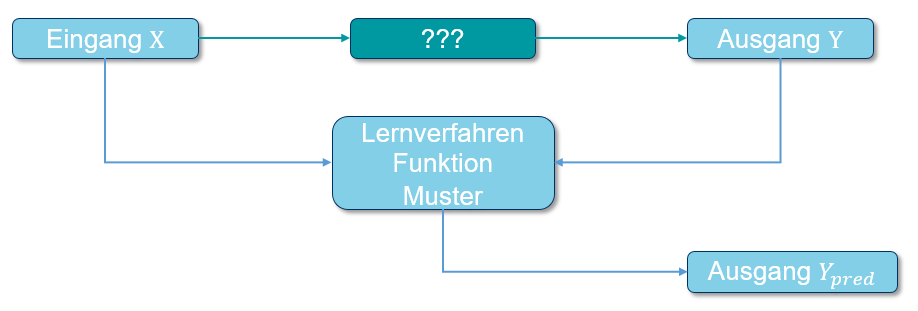


Abb. 3: Maschinelles Lernen

● Überwachtes Lernen (Supervised Learning)

Ziel besteht darin, dass eine Abbildung von Eingabewerte auf Ausgabewerte gelernt werden soll, wobei die korrekten Ausgabewerte für die dazugehörten Eingabewerte im Datensatz bekannt sind. Die Ergebnisse des Lernprozesses können mit den bekannten richtigen Ergebnissen verglichen, also „überwacht“, werden.

In der Arbeit sind alle Lernprozesse überwachtes Lernen, weil die Zielgrößen (maximale Verschiebungen) schon im Trainingsdatensatz bekannt sind.

● Unüberwachtes Lernen (Unsupervised Learning)

Im Gegensatz zum überwachten Lernen sind die Ausgabewerte im Voraus nicht bekannt und uns stehen nur die Eingabewerte zur Verfügung. Der Lernprozess versucht, in den Eingabedaten Muster und Regelmäßigkeiten zu erkennen. In anderen Worten, beim unüberwachten Lernen soll die in Eingangsdaten versteckten Ähnlichkeiten herausgefunden und mit einem mathematischen Modell beschreibt werden soll.

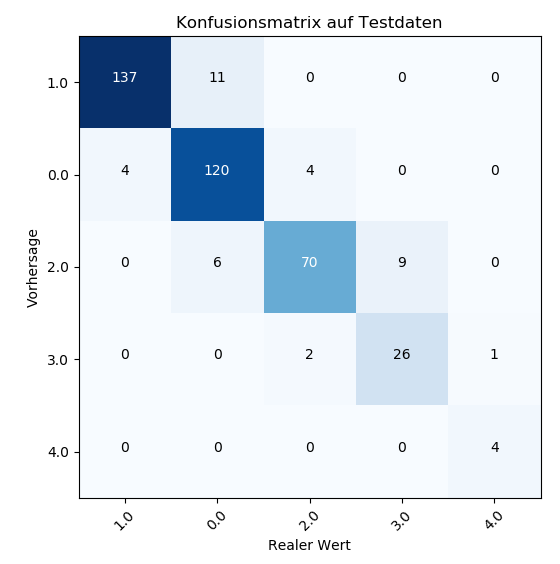
● Verstärkendes Lernen (Reinforment Learning)

Gleich wie die beim unüberwachten Lernen uns stehen keine Ausgabewerte zur Verfügung, aber beim verstärkenden Lernen enthält das Algorithmus mit jedem Beispiel positives bzw. negatives Feedback zu seiner vorschlagenden Lösung. Das Feedback heißt also Belohnung. Der Algorithmus heißt also Agent. Ein Agent entscheidet selbstständig, welche Aktion in welcher Situation die beste ist, womit die beste Belohnung erzielen kann.

● Klassifikation – Modell zum Vorhersage der Klassenzugehörigkeit

Klassifikationsverfahren sind Methoden zur Klassifizierung von Objekten oder Situationen in Klassen. Mithilfe der Klassifikation werden Objekte oder Datensätze einer vorgegebenen Anzahl von Klassen zugeordnet. Hierzu wird ein Modell auf Grundlage von Datensätzen erzeugt, deren Klassenzugehörigkeit bekannt ist. D.h. das Klassifikationsverfahren gehört zum überwachten Lernen.

Bei der Klassifikation werden Objekte anhand von bestimmten Merkmalen durch ein Klassifikationsmodell bzw. Klassifikator in verschiedene Klassen eingeordnet. Dabei macht der Klassifikator eigentlich auch Fehler. Zur Beurteilung eines Klassifikators kommt die Konfusionsmatrix zum Einsatz.



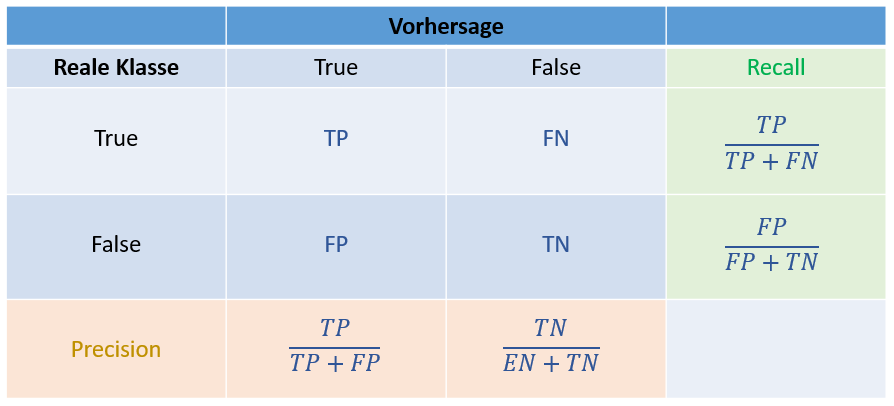


Abb. 4: Konfusionsmatrix

Abb. 4 oben zeigt ein beispielsweise Konfusionsmatrix. Davon können unterschiedliche Kennwerte, die die Modellgüte beschreiben, ausgelesen werden. Diese Kennwerte zeigt Abb.4 unten, nämlich Recall und Precision. Für eine vereinfachte Beurteilung eines Klassifikators nutzt man im Praxis i.d.R. die Genauigkeit, die Formel lautet:

|  |  |
| --- | --- |
|  | (1) |

Die Genauigkeit kann sich auf mehr Klassen erweitern.

● Regression – Modell zum Vorhersage der stetigen Werten

Im Gegensatz zur Klassifikation nimmt die Zielvariable stetige Zahlwerte an. Trotzdem stehen die richtigen realen Zielvariable auch im Datensatz zur Verfügung. Daher zählt Regression zum überwachten Lernen. Mithilfe der Regression wird ein funktionaler Zusammenhang zwischen Zielvariable und Einflussvariable eines Datensatzes hergestellt.

Dies entspricht die Anforderung in der Arbeit, die maximalen Verschiebungen (stetige Zahlwerte) vorherzusagen. Deswegen ist Regression einen wesentlichen Teil des Vorhersagemodells in der Arbeit.

Zur Charakterisierung der Modellgüte eines Regression-Vorhersagemodells kommen einige Kennwerte zum Einsatz, nämlich MAE, MAPE, MSE usw.

MAE, auf Engl. mean absolut error, die Formel davon lautet:

|  |  |
| --- | --- |
|  | (2) |

wobei folgende Variablen Verwendung finden:

: Anzahl der Vorhersagewerte

: Vorhersagewert

: Realer Wert

Wegen der Absolutwert in der Formel beschreibt MAE lediglich die Höhe der Abweichung der Vorhersage von der realen Werten, nicht jedoch die Abweichungsrichtung (positive oder negative Abweichung). Außerdem ist MAE außer dem mittleren Wertbereich nicht aussagekräftig, z.B. für , hat das Vorhersagemodell größere Abweichung, aber für , hat das Modell ausgezeichnete Güte.

Daher kommt MAPE, auf Engl. mean absolut percentage error. Die Formel davon lautet:

|  |  |
| --- | --- |
|  | (3) |

wobei ist die Notation ebenso wie Gl. (2).

Im Vergleich zu MAE ist jede Abweichung bezogen auf die entsprechende reale Werte. Deswegen besitzt MAPE in ganzen Wertbereich gleichmäßige Aussagekraft. Wäre ein 0 Wert ) im Datensatz, würde MAPE nicht funktionieren, weil die im Nenner in Gl. (3) steht. In der Arbeit tritt diese Situation nicht ein, weil sowohl die maximale Spannung als auch die maximale Verschiebung eines unter Lasten beauftragten Bauteils nicht auf 0 sein können.

MSE, auf Engl. mean squared error, die Formel davon lautet:

|  |  |
| --- | --- |
|  | (4) |

Die ist sehr ähnlich wie MAE, der Unterschied liegt daran, dass die Streuung der Vorhersage mit quadratischer Differenz charakterisiert ist. Da besitzt MSE ebenso die Vor- und Nachteile von MAE.

● Clustering

Clustering ist ähnlich zur Klassifikation, die Beispiele in Klassen bzw. Gruppen einzuordnen. Der Unterschied liegt daran, dass bei Clustering die Klassenzugehörigkeit (Zielgrößen) nicht im Datensatz bekannt sind. Dies entspricht genau die Definition des unüberwachten Lernen. Diese Verfahren haben die Aufgabe, verständliche Muster und Zusammenhänge zur Beschreibung eines komplexen Datensatzes zu generieren, ohne dass vorher bekannt ist, welche Zusammenhänge innerhalbe der Daten existieren [8].

● Assoziation

Die Assoziation bezeichnet die Suche nach Abhängigkeit zwischen unterschiedlichen Merkmalswerten innerhalb eines Datensatzes. In anderen Worten heißt es, die relevanten Korrelationen zwischen Variablen in großen Datensatz aufzudecken.

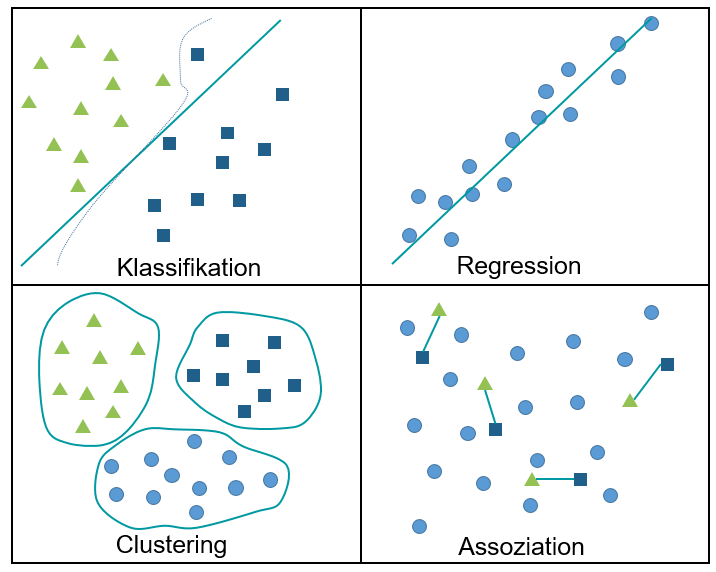


Abb. 5: Zielstellungen des maschinellen Lernens

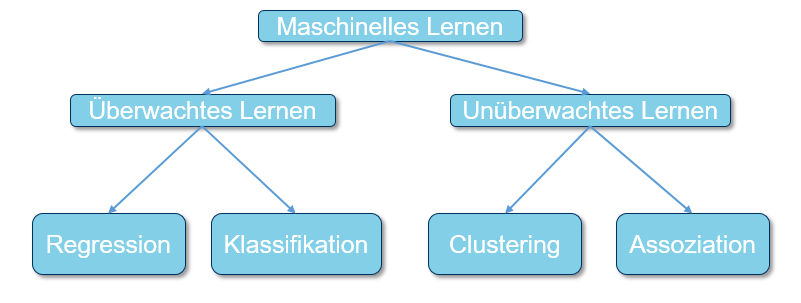


Abb. 6: Gliederung der Definitionen von maschinellen Lernen

2.1.2 Gradientenabstiegsverfahren und Verallgemeinerung

Zur Charakterisierung der Abweichung von Vorhersage und Richtige ist die Kostenfunktion (Beispiel in Kap. 6.1). Ziel des Trainierens ist die Kostenfunktion zu minimieren. Dazu braucht man Gradientenabstiegsverfahren, wobei geht die Kostenfunktion in Richtung des negativen Gradienten ab, bis keine numerische Verbesserung mehr erzielt werden kann. Das Verfahren lautet:

|  |  |
| --- | --- |
|  | (5) |

Die entsprechende grafische Darstellung zeigt die Abb. 7. Davon geht die Kostenfunktion schrittweise zum Minimum, wobei wird im jeden Schritt die Gewichte aktualisiert.

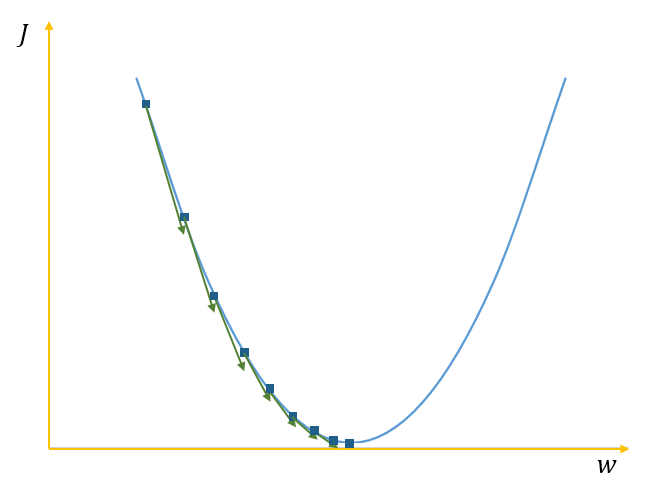


Abb. 7: Gradientenabstiegsverfahren

Ein Vorhersagemodell, das nicht ganz richtig verallgemeinern kann, ist entweder Overfitting oder Underfitting. Eine genaue Anpassung ermöglicht es, das Modell nach der Lernphase richtig zu verallgemeinern.

Beim Overfitting eines Modells heißt es also, dass das Modell die Datenpunkte überangepasst ist. Eine zu komplexe Modell erlernt nicht nur die zugrundliegende Funktion, sondern auch Rauschen in den Daten (siehe Abb. 8 links).

Im Gegensatz dazu ist ein Modell beim Underfitting weniger komplex als die zugrundeliegende Funktion, welche damit nicht vollständig beschreibt werden kann. Z.B. Zielgröße hängt quadratisch von Einflussgrößen ab, aber es wurde nur ein lineares Modell gewählt (siehe Abb. 8 rechts).

Eine einfache Methode im Praxis die Anzahl des Neurons und der Schichten in einem neuronalen Netz auszuwählen ist „stretchhose“ (auf engl. Stretch pants, instead of wasting time looking for pants that perfectly match your size, just use large stretch pants that will shrink down to the right size.), wobei wird ein Modell mit mehr Schichten und Neuronen als es eigentlich braucht, dann lässt sich Regularizationstechnik (z.B. early stopping, dropout) einsetzen, das Auftreten von Overfitting zu vermeiden [9].

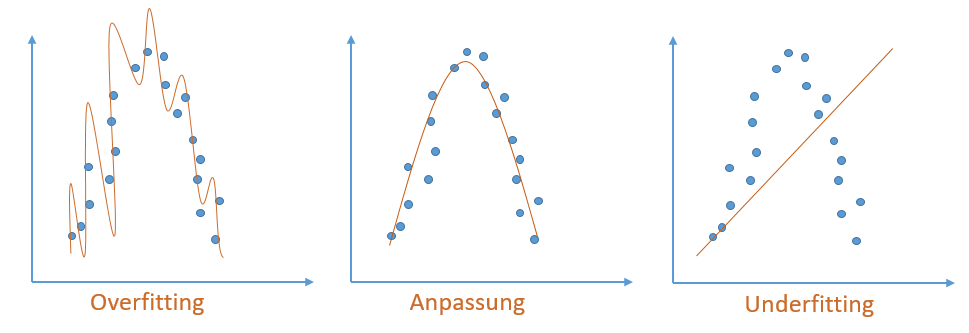


Abb. 8: (Links) Overfitting, (Rechts) Underfitting

## 2.2 Methoden im maschinellen Lernen

Es gibt umfangreiche Methoden im maschinellen Lernen. Die enthalten lineare Regression, LASSO, Support Vector Maschine (SVM), Entscheidungsbaum, Neuronales Netz (NN) usw. Je nach den Eigenschaften des Problems wird dazu geeignete Methode eingesetzt. D.h. man muss zunächst die umfangreichen Methoden bekannt sein, in welcher Situation eine angepasste Methode eingesetzt werden könnte. Nach der Abwägung von Komplex und Reichenleistung werden SVM und NN in der Arbeit genommen.

2.2.1 Support Vector Maschine (SVM)

Eigentlich ist SVM ein 2-Klassen-Klassifikator (zu mehren Klassen erweiterbar), dessen Strategie zum Lernen ist der Abstand (auf Englisch Margin) zwischen die zwei Klassen im Featureraum zu maximieren.

Abb. 10 zeigt ein Beispiel von SVM. Die Trennebene verhält sich als Kriterium zur Klassifizierung der Datenpunkte.

Falls , dann Zuordnung zu positiven Klasse, Zielgröße .

Falls , dann Zuordnung zu negativen Klasse, Zielgröße .

Im Nachfolgendes wird die mathematischen Hintergründe dargestellt, wodurch wird der max. Margin erreicht [10].

● Funktioneller Margin

Der Absolutwert ist der relative Abstand zwischen Datenpunkt und Trennebene\*. Weiterhin ergibt sich den gleichen Zahlenwert wie den Absolutwert.

|  |  |
| --- | --- |
|  | (6) |

Damit kommt die Definition von funktionellen Margin :

|  |  |
| --- | --- |
|  | (7) |

wobei, ist ein Datenpunkt im Featureraum, ist Label im Datensatz.

Weiterhin heißt ) der funktionelle Margin einer Trennebene bezogen auf vorgegebenen Datensatz.

|  |  |
| --- | --- |
|  | (8) |

Falls die Koeffizienten sich verdoppeln würden, würde die Trennebene sich nicht bewegen aber der funktionelle Margin sich verdoppeln. In andren Worten reicht es nicht, den Abstand vollständig zu beschreiben.

● Geometrischer Margin

Daher kommt die Definition von geometrischen Margin.

|  |  |
| --- | --- |
|  | (9) |

wobei, ist 2-Norm eines Vektors .

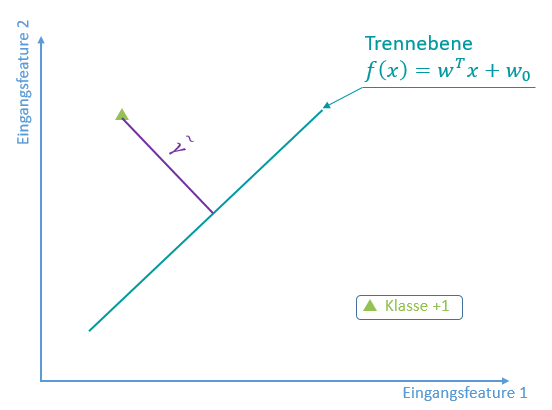


Abb. 9: Geometrischer Margin

Einen Absolutwert zu erwerben wird nochmal zu Gl. (9) multipliziert. Danach ergibt der vollständige geometrische Margin durch:

|  |  |
| --- | --- |
|  | (10) |

Davon entspricht der Zählerpolynom zum funktionellen Margin .

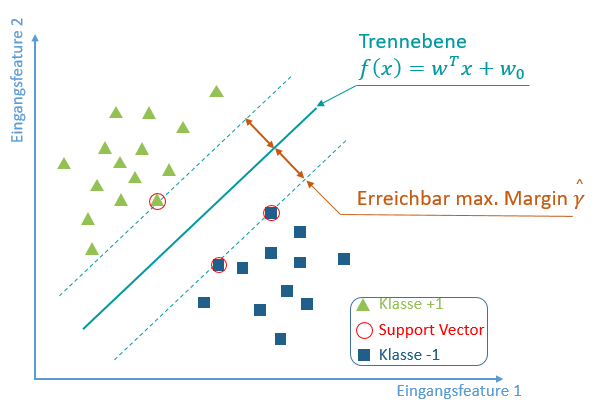


Abb. 10: Support Vector Maschine

Wie in Abb. 10 gezeigt, Ziel von uns ist es, der geometrische Margin zu maximieren:

|  |  |
| --- | --- |
|  | (11) |

unter Bedingung (Definition vom funktionellen Margin ):

|  |  |
| --- | --- |
|  | (12) |

Annahme:

|  |  |
| --- | --- |
|  | (13) |

Dann Gl. (11) ändert sich zu:

|  |  |
| --- | --- |
|  | (14) |

\*Grund der Annahme: weil die Koeffizienten proportional sich verändern können und damit bewegt die Trennebene nicht, ist die Annahme logisch.

Damit ist der Prozess, wodurch Gl. (14) gezeigt, die Suche nach der besten Trennebene. Die Datenpunkte, die genau in Abb.10 auf den gestrichelten Linien liegen, heißen als Support Vector.

● Kernel Funktion

Das obengenannte Verfahren besitzt eine Voraussetzung, dass die Datenpunkte im Featureraum linear trennbar sein müssen, sonst weißt es eine schlechte Modellgüte beim Einsatz auf. Eine Möglichkeit dies zu lösen ist es, das nicht linear trennbare Datensatz in einen höherdimensionalen Raum zu transformieren, in welchem man sich eine bessere lineare Trennbarkeit erhofft.

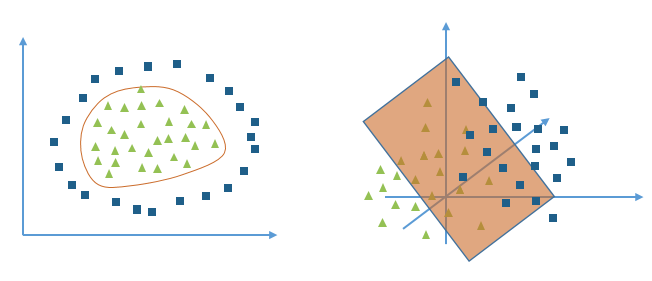


Abb. 11: Datensatz in höherdimensionalen Raum transformieren

Abb. 11 zeigt das Transformationsprozess, wobei mittels Kernel Funktion ein Datensatz sich in linear trennbare umwandeln lässt. Die üblichen Kernels sind:

lineare Kernel,

|  |  |
| --- | --- |
|  | (15) |

polynomiale Kernel,

|  |  |
| --- | --- |
|  | (16) |

wobei einer ganzen Zahl.

RBF-Kernel,

|  |  |
| --- | --- |
|  | (17) |

● Soft Margin

Beim Soft Margin heißt es, dass SVM eine Toleranz für Ausreißer hat, somit kann ein größere Margin erzielen. Dazu wird die Nebenbedingung in Gl. (14) erweitert:

|  |  |
| --- | --- |
|  | (18) |

Steuerung dieses weichen Fehlers durch Hyperparameter in der Kostenfunktion:

|  |  |
| --- | --- |
|  | (19) |

Ist groß, so ergeben sich eher enge Margin, die eher selten verletzt werden, d.h. der Klassifikator passt sich sehr stark an die Daten an. Ist klein, so ergeben sich eher breitere Margins, die öfter verletzt werden, d.h. der Klassifikator passt sich den Daten nicht im Detail an.

Mit Strafterm Kompromiss zwischen Komplexität (Anzahl der Stützvektoren) und Datenfehlanpassung (Anzahl an nicht trennbaren Punkten).

2.2.2 Neuronales Netz

● Perzeptron

Das Perzeptron ist eine von einfachsten künstlichen neuronalen Netzen. Es wurde in 1957 von Frank Rosenblatt aufgestellt. Die Knotenwerte der eingehenden Schicht werden gewichtet und anschließend aufsummiert. Zum Schluss lässt die Summe sich mittels Aktivierungsfunktion transformieren, eine Zielgröße vorherzusagen. Abb. 12 ist die schematische Darstellung dieser Funktionsweise. Davon Einflussgröße , Zielgröße , Gewichte , Aktivierungsfunktion

|  |  |
| --- | --- |
|  | (20) |

Falls , dann Zuordnung zu positiven Klasse, Zielgröße .

Falls , dann Zuordnung zu negativen Klasse, Zielgröße

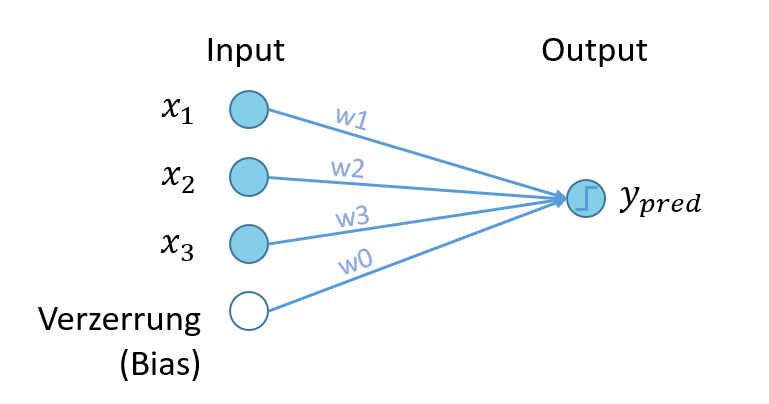


Abb. 12: Perzeptron

Trainieren eines Perzeptrons ist es so, die richtigen Zahlenwerte für die Gewichte herauszufinden. Es fangt mit bestimmten Initialwerten an Gewichte an, dann nach jeder Iteration aktualisieren sich die Gewichte, um die richtigen Zielgrößen nachzufolgen. Die Iteration könnte beendet werden, wenn die hinreichende Genauigkeit erreicht.

Zur Charakterisierung der Abweichung zwischen Richtige und Vorhersage kommt eine Kostenfunktion im Einsatz. Je nach den Arten und Zielstellungen eines maschinellen Lernens ist das Formel der Kostenfunktion unterschiedlich. Ziel des Trainierens im Allgemeinen ist es, die Kostenfunktion sich möglichsten verkleinern zu lassen.

Die Gewichte ergeben sich wieder durch eine Minimierung der Fehler zwischen Richtige und Vorhersage, also der Kostenfunktion des Perzeptrons, für ein Trainingsbeispiel mit Klasse und Vorhersage anhand der Kreuzentropie:

|  |  |
| --- | --- |
|  | (21) |

Fehler- bzw. Kostenminimierung wird mittels Gradientenabstiegsverfahren erledigt, damit die Gewichteupdate durch:

|  |  |
| --- | --- |
|  | (22) |

wobei

|  |  |
| --- | --- |
|  | (23) |

mit der Lernrate.

● Mehrlagige Schichten von Perzeptronen

Die mehrlagigen Schichten von Perzeptronen heißen also das sogenannte neuronale Netz. Außer den Ein- und Ausgangsschichten sind die dazwischenliegenden Schichten von Perzeptronen Zwischenschichten (auf Engl. Hidden Layer). Je mehr Zwischenschichten es gibt, desto tiefer ist das neuronale Netz, im Englischen spricht man daher auch von Deep Learning.

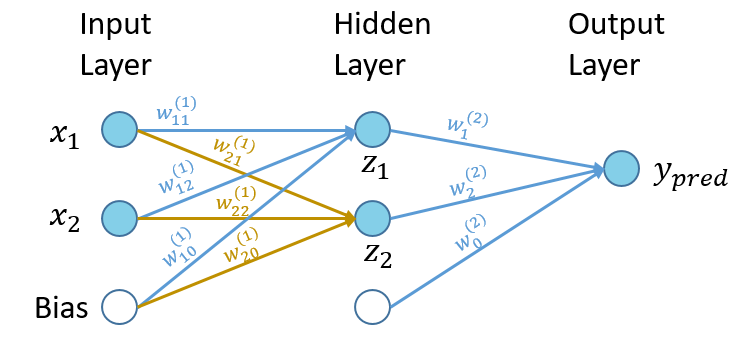


Abb. 13: Neuronales Netz

Abb. 13 zeigt ein Beispiel vom neuronalen Netz. Jede Verbindung besitzt ein eigenes Gewicht. Auf Grund der komplexen Strukturen wird die Notation für das Gewicht auch erweitert, nämlich:

Jede Einheit der Zwischenschichten berechnet sich wie das Perzeptron mit einer Aktivierungsfunktion . Damit könnte Output sich durch die Gl. (24) ergeben:

|  |  |
| --- | --- |
|  | (23) |

mit , der Aktivierungsfunktion jeweils für Hidden Layer und Output Layer. Typische Aktivierungsfunktionen sind logistische Funktion, rectified linear unit (ReLU), exponential linear unit (ELU) usw. Bei den Anwendungsfällen werden unterschiedliche Ergebnisse geliefert, wenn dieselbe neuronale Netz mit unterschiedlicher Aktivierungsfunktion verseht ist. Aber wie Anrelion Georon in seinem Buch „Hands-On Maschine Learning with Scikit-Learn and Tensorflow“ gesagt: ELU ist normalerweise besser als die andere [9].

In Kap. 2.1.1 sind die Zielstellungen des maschinellen Lernens vorgestellt, davon uns interessieren in der Arbeit Regression und Klassifikation. Die neuronalen Netze sind in der Lage, sowohl Klassifikation als auch Regression zu erledigen. Ob ein neuronales Netz für Klassifikation oder Regression ist, hängt es im Wesentlichen von der Aktivierungsfunktion des Output Layers ab. Für ein Klassifikationsproblem wird normalerweise eine logistische Funktion als Aktivierungsfunktion im Output Layer eingesetzt, welche die Aufsummierung der Verbindungen zu einer Wahrscheinlichkeit umrechnen kann. Damit wird die Klasse, die über höchster Wahrscheinlichkeit verfügt, als die vorhergesagte Klasse angesehen. Im Gegensatz dazu sind die Zielgrößen einer Regression stetige Zahlwerte. D.h. eine Aktivierungsfunktion im Output Layer bei einer Regression ist unnötig, wobei eine Aufsummierung der Verbindungen reicht, eine stetige Zielgröße bereitzustellen.

● Kostenfunktion

Die Definition der Kostenfunktion und deren Funktionsweise sind schon in der Vorstellung des Perzeptrons dargestellt. Für neuronales Netz ist die Situation mehr komplex als die für ein einzelne Perzeptron.

Anhand des Beispiels eines neuronalen Netzes in Abb. 13 sich unterscheiden die Kostenfunktion für Klassifikation und Regression, nämlich:

für Regression, quadratischen Abweichung für ein Datenpunkt mit Vorhersage :

|  |  |
| --- | --- |
|  | (24) |

für Klassifikation, kreuzentropie für ein Datenpunkt mit Vorhersage :

|  |  |
| --- | --- |
|  | (25) |

Aktivierung der Gewichte und Berechnung der Aktivierung ergeben sich durch:

|  |  |
| --- | --- |
|  | (26) |
|  | (27) |

wobei der Schichtnummer.

Nutze Kettenregel, um Einfluss der Gewichte am Anfang des Netzes zu bestimmen:

|  |  |
| --- | --- |
|  | (28) |
|  | (29) |
|  | (30) |

Der Einsatz von Kettenregel wird beim Trainieren eines neuronalen Netzes Backpropagation genannt, nämlich zwei Phasen: Vor- und Rückwärtsprozess [11].

Beim Vorwärtsprozess lässt sich ein Datenbeispiel ins Netz hereingehen, dadurch werden die Ergebnisse vom allen Neuronen ausgerechnet inklusive der Vorhersage von Output Layer. Danach wird die Vorhersage mit Richtige verglichen, die Abweichung lässt sich mit Kostenfunktion interpretieren.

Beim Rückwärtsprozess fließt die Abweichung von Output Layer zum Input Layer durch, wobei die Einflüsse von allen Gewichten der Verbindungen quantifiziert werden.

Zum Schluss lassen sich die Gewichte von allen Verbindungen mittels Gradientenabstiegsverfahren so verändern, die Abweichung zu minimieren.

# 3. Generierung der FEM-basierten Daten

## 3.1 Vorgehensweise einer FEM-Analyse

Die Methode der finiten Elemente (FEM) ist ein durch numerische Iterationen berechnetes Verfahren, das in vielen Anwendungsbereiche des Maschinenbaus und Elektromagnets zum Einsatz kommt. Die Grundgleichungen zur Beschreibung strukturmechanischer Probleme wie Deformationen, Spannungen, Geschwindigkeiten, Druck, Temperaturen usw., sind gewöhnlich oder partielle Differenzialgleichungen (DGLn) bzw. Differenzialgleichungssysteme [12].

Der Grundgedanke von FEM besteht darin, dass die uns interessierende Bereiche in eine endliche Anzahl einfacher Teilbereiche. In FEM wird dieser Prozess Diskretisierung genannt, in der die großen Bereiche zu zerkleinern. Die Differenzialgleichungen, die das physikalische und mechanische System beschreibt, werden auf der endlichen Anzahl von Elementen gelöst.

Im Ingenieurbereich stehen vielseitige Simulationssoftwaren zur Verfügung, die eine FEM-Analyse durchführen zu können. Z.B. ABAQUS, ANSYS und SolidWorks (mit Simulation-AddIn). Obwohl es viele Auswahl im Markt gibt, haben alle sehr ähnliche Vorgehensweise. Die beispielsweise in SolidWorks-Simulation können in 7 Schritte eingegliedert [13].

● Erstellen einer Studie

In der SolidWorks-Simulation Zusatzanwendung stehen nicht nur konventionale lineare Mechanikanalyse, sondern auch mehre Einsatztype zur Verfügung, z.B. Thermische Analyse, Frequenzanalyse, Knickenanalyse usw. (siehe Abb. 14).

In der Arbeit wird auf Festigkeitsanalyse (Static in Abb. 14) konzentriert.



Abb. 14: Einsatztype in SolidWorks Simulation

● Anwenden des Materials

Einige am häufigsten eingesetzte Materialien sind in der Bibliothek vordefiniert und mit dazu entsprechend Name gekennzeichnet. Die benötigten Kennwerte von Werkstoff, z.B. Elastizitätsmodul und Schubmodul, sind auch dazu verbunden.

In der Arbeit wird nur ein Material „AISI 1020“ (DIN C22) als Beispiel untersucht.

● Einspannungen definieren

Ziel davon ist die Bewegungen eines Bauteils im Raum ein oder mehr Freiheitsgrad einzuschränken. Im Allgemeine heißt es auch Randbedingungen Definieren. Der Grund liegt daran, dass die Differenzialgleichungen der Mechanik die Randbedingungen benötigen, um gelöst werden zu ermöglichen. Die Type der Einspannung ist je nach dem einzelnen Szenario der Anwendung.

In der Arbeit wird eine Stirnfläche als fixiert betrachtet.

● Lasten definieren

Im Ingenieurbereich gibt’s unterschiedliche Lasten. Davon sind die äußeren Kräfte und Momente als allgemeine Lasten betrachtet.

In der Arbeit sind nur Kräfte als bestimmte Lasten berücksichtigt.

● Modell vernetzen

Im Rahmen der FEM ist die Vernetzung eines Bauteils entscheidend. Damit entstehen diese endlichen Elemente. Je nach Form, Größe und Anzahl der Elemente können die entsprechenden Ergebnisse sich voneinander abweichen.

Theoretisch stehen in der FEM vielseitige Elementtypklasse zur Verfügung, nämlich für Linienelemente gibt’s Stab und Balken, für Flächenelemente gibt’s Dreieck und Viereck, für Volumenelemente gibt’s Hexaeder und Tetraeder.

Bei SolidWorks Simulation sind drei grundlegende Vernetzungstype verfügbar: shell mesh, beam mesh und solid mesh [14]. Folgendes steht die Eigenschaft der typischen Vernetzungstype (siehe Tabelle 1).

Tabelle 1: Vergleich der Vernetzungstype bei SolidWorks Simulation

|  |  |  |  |
| --- | --- | --- | --- |
|  | Solid Mesh | Shell Mesh | Beam Mesh |
| Elementtype | Tetraeder | Dreieck | Balken |
| Einsätze | Für alles andere  (als standardmäßige Voreinstellung) | Dünne Bauteile | Bauteile mit konstanter Abschnitt |

Für Solid Mesh bei SolidWorks Simulation bittet der Lieferant keine andere Elementtype, sondern nur Tetraeder. D.h. in der Arbeit spielt der Elementtyp keine Rolle.

Die Größe und die Anzahl der Elemente sind sich voneinander abhängig. Je kleiner die Elemente sind, desto größere Anzahl der Elemente ist.

In der Abb. 15 wird Einfluss der Elementsgröße auf FEM Ergebnisse (maximale Verschiebung) gezeigt. Elementsgröße 0,2mm ist als angenommene Richtige bzw. Referenz betrachtet. Es weist eine Tendenz auf, dass mit absteigenden Elementsgröße neigt es dazu, größere Ergebnisse zu erzielen. Im Vergleich zu großen Elementsgröße besitzen die Vernetzungen, die mit kleinen Elementsgröße versehen sind, einen deutlichen Unterschied zur Referenz, dazu zählen sowohl die Verteilung als auch die Differenz.

In der Arbeit wird der Unterschied nicht weiter qualitativ untersucht, sondern nehmen wir eine quasi vernünftige Elementsgröße 0.2mm als Beispiel. Die ist nach der Abwägung zwischen Genauigkeit und Rechenzeit.



Abb. 15: Einfluss der Elementsgröße auf FEM Ergebnisse

(maximale Verschiebungen)

● Studie ausführen

Wenn alle benötigten Eingangsdaten (Geometrie, Material, Lasten, Vernetzung usw.) zugewiesen sind, dann kann das Berechnungsverfahren selbst automatisch durchlaufen.

● Ergebnisse analysieren

Beim Standard stehen 3 verformungsrelevante Ergebnisse zur Verfügung, nämlich vonMises Spannungen, resultierende Verschiebungen und äquivalente Dehnungen.

In der Arbeit uns interessieren die vonMises Spannungen und resultierende Verschiebungen.

## 3.2 Automatisierung der Vorgehensweise durch SolidWorks-API mittels C# Programmierung

C# ist eine objektorientierte Programmierungssprache, die erste Version C# 1.0 war in 2002 von Microsoft veröffentlicht. Nach mehr als zehnjährige Entwicklungszeit hat C# große Fortschritte gemacht. Die mit der C# entwickelte Programme laufen auf der Microsoft .NET-Plattform. Dabei handelt es sich um eine virtuelle Maschine, die Programmen, die für diese Plattform entwickelt wurden, einen Prozessor vorgaukelt, der so in Form eines echten Computerchips gar nicht existiert [15].

In der Arbeit wird die Version C# 7.5.2 eingesetzt.

API heißt auch Programmierschnittstelle, mit der von einem Softwaresystem die anderen Programme zur Anbindung an das System zur Verfügung gestellt wird. Es gibt zwei Type von API [16]. Eine ist mit Abhängigkeit von Programmierungssprache, damit können sowohl die Syntax als auch die Elemente von dieser bestimmten Programmierungssprache zugegriffen werden. Im Gegensatz ist die andere ohne Abhängigkeit von Programmierungssprache. D.h. die kann von mehreren Programmierungssprachen angerufen werden [17].

SolidWorks ist ein kommerzielles CAD System von Dassault Systemes. Das umfassen vollständige 3D-Modellierung und umfangreiche zusätzliche Anwendungen (auf Englisch Add-In-Module), SolidWorks Simulation ist eine davon. Neben dem bietet Dassault Systemes SolidWorks-APIs in C#, VB und C++ für die Automatisierung und Anpassung der Vorgehensweise. In der Arbeit wird C# als API-Programmierungssprache eingesetzt.

Im SolidWorks gibt’s drei grundlegende Dokumente, nämlich Bauteil, Komponente und Zeichnungen. Entsprechend stehen Drei dazugehörten Objekte in SolidWorks-API, nämlich PartDoc, AssemblyDoc und DrawingDoc. Die Struktur der Objekte in SolidWorks API ist in Abb. 16 dargestellt. SldWorks funktioniert als Wurzel und alle sind davon deriviert. Jede Objekte hat seine eigenen Funktionen, damit die Manipulationen in hinterlegten Applikationen erledigt werden. Darauffolgende ist ein Beispiel.

„*swModel.SketchManager.CreateLine(0, 0, 0, 1, 0, 0);*“

Eine Instanz von ModelDoc2 heißt „*swModel*“. SketechManager ist eine von dazugehörten Objekten. Die fasst alle Manipulationen um, damit der Sketch erstellt u./o. ergänzt werden können. CreateLine ist eine Funktion davon, die eine Linie anhand der vorgegebenen Koordinaten im Sketch hinzufügen kann.

In der Arbeit wird SolidWorks 2016 und dazu entsprechende API verwendet.



Abb. 16: Struktur der Objekte in SolidWorks API

Für das Training eines wissensbasierten Modells (z.B. Neuronales Netz) ist i.d.R. eine hohe Anzahl von Daten unerlässlich. Bei manchen Anwendungen wie Bilderarbeitung und Spracheerkennung können die benötigten Daten im Internet herausgefunden und heruntergeladen werden. Aber bei dieser Arbeit steht keine bestimmten bereiten Daten zur Verfügung. Trotzdem ist es tausende Daten manuell anzusammeln unrealistisch. Deswegen spielt die Automatisierung der Vorgehensweise eine große Rolle.

Je nach der Geometrie und Lasten unterscheidet die Anzahl des Eingangsfeatures (siehe Tabelle 2) sich. Hauptsache ist es mit dieser Eingangsfeature der Bauteil und dessen äußeren Lasten vollständig definiert werden zu können. Tabelle 2 zeiget die Beispiele (siehe Abb. 1) und dazugehörte Anzahl des Eingangsfeatures.

Tabelle 2: Anzahl des Eingangsfeatures bei unterschiedlichen Szenen

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Balken | L-Profil | W-Profil | W-Profil  (Planare Deformation) |
| Anzahl des Eingangsfeatures | 6 | 7 | 9 | 8 |

Für eine gute Verallgemeinerung eines Vorhersagemodells sollen die gefütterte Traningsdaten möglichst zufällig sein. D.h. die Daten dürfen keine subjektiven Präferenzen enthalten. In C# ist Random-Funktion eine der Möglichkeiten, Zufallszahl zu erzeugen. Man muss darauf beachten, dass Random-Funktion in C# keine mathematisch echte Zufallszahl erzeugen kann, sondern Pseudozufallszahl.

In der Arbeit wird die Pseudozufallszahl verwendet. Der Unterschied der Wirkung auf Vorhersagemodell zwischen Pseudozufallszahl und mathematisch echten Zufallszahl wurde nicht tiefer untergesucht.

Das Wertebereich von Eingangsfeature beschränkt sich innerhalb einer sinnvollen Grenz. Der Grund liegt daran, dass die Komplexität des Vorhersagemodells sich auch in einem gewissen Niveau befindet. Dazu entsprechen es ein relativ kleiner Umfang von Trainingsdatensatz und eine quasi niedrige Rechenleistung. In Tabelle 3 steht das Wertbereich des Eingangsfeatures von bespielweise W-Profil-Bauteile.

Tabelle 3: Wertbereich des Eingangsfeatures von W-Profil-Bauteile

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | m | width | thickness | L1 | L2 | L3 | F1 | F2 | F3 |
| Wert-bereich | [2, 4] | [3, 10] | [1, 10] | [24, 32] | [15, 21] | [14, 22] | [0, 50] | [0, 50] | [0, 50] |

Die manuelle Vorgehensweise einer FEM-Studie ist schon in Kapitel 3.1 gezeigt. Bevor jeder räumlichen FEM-Studie muss 3D-Modell vorhanden sein. D.h. zuerst muss man ein geometrisches 3D-Modell anhand der durch die Random-Funktion generierten Eingangsdaten erstellen. Danach läuft eine entsprechende FEM-Studie durch. Um vollständige automatische Durchführungen zu realisieren, wird eine for-Schleife im Programm geschrieben. Der konkrete C# Quellecode wird im Anhang beigefügt.

Am Ende jeder for-Schleife können alle FEM-Ergebnisse durch dazu entsprechende API-Funktion abgelesen werden. Ziel der Arbeit ist es die maximale Verschiebung eines Bauteils unter äußeren Lasten vorherzusagen, deswegen ist die maximale Verschiebung im Trainingsdatensatz unerlässlich. Darüber hinaus funktioniert die entsprechende maximale Spannung als ein zusätzlicher Faktor, der sich um die Streckgrenze behandelt (siehe Kapitel 6.3). Nach erfolgreicher Ablesen von FEM-Ergebnissen wird die Zielgröße durch Microsoft-Excel-API in einer Tabelle (.xlsx) exportiert.



Abb. 17: Blockdiagramm der Datenstruktur



Abb. 18: Aufbau der Datenstrukturen von W-Profil

(räumliche Deformation)

I.d.R. ist ein Datensatz wie Abb. 17 von vier Teilen aufgebaut, nämlich Geometrie, Lasten, Zielgröße und Zusatz (wenn nötig). In der Arbeit wird eine Kombination von die 4 (o. 3) Eigenschaften eine Datei genannt. Z.B. eine Zeile in Abb.18 heißt eine Datei.

In Abb. 18 wird Aufbau der Datenstrukturen von W-Profil mit räumlicher Deformation dargestellt. Davon sind die zwei „maxStress(MPa)“ und „maxDisp(mm)“ die obere genannte zwei Zielgröße. Rechts liegen die durch Zielgröße abgeleitet Eigenschaften, „class“ und „out“, die als Zusatz betrachtet werden.

# 4 Statistische Untersuchung

## 4.1 Korrelation innerhalb des Datensatz

Nach der automatisierten Datenerfassung steht eine Tabelle zur Verfügung, die alle Eingangsparameter und Zielgröße enthaltet. Beim Erkunden von Daten geht es um Statistik. D.h. in diesem Schritt werden alle Eigenschaften, die nicht direkt entdeckt werden können, mit Hilfe der statistischen Beschreibungen einfach und explizit interpoliert. Dazu zählen sowohl die Verteilung als auch die sich miteinander verbundenen Korrelationsindex. Ziel davon ist es, Universalität von Daten zu überprüfen, die bei der Verallgemeinerung des Vorhersagemodells eine entscheidende Rolle spielt.

Eine Korrelation beschreibt eine Beziehung zwischen zwei oder mehreren Merkmalen, Ereignissen, Zuständen oder Funktionen [18]. Die Maßzahlen der Korrelation liegen betragsmäßig meist in einem Bereich von -1 (ein vollständig negativer linearer Zusammenhang) bis 1 (ein vollständig positiver linearer Zusammenhang). Für zwei quadratisch integrierbare Zufallsvariablen *X* und *Y* mit jeweils positiver Standardabweichung bzw. und Kovarianz Cov(*X*, *Y*) ist der Korrelationskoeffizient (Pearsonscher Maßkorrelationskoeffizient) definiert durch:

|  |  |
| --- | --- |
|  | (31) |

In Abb. 19 wird die Korrelation des Eingangsfestures punktweise dargestellt. Die Diagonale steht die sogenannte Kerndichteschätzung, die wird im Kapitel 4.2 erläutert. Außer der Diagonale sind die andere Zelle paarweise Diagramm, in dem wird ein Feature zu alle anderen punktweise aufgezeichnet. Die entsprechenden pearsonscher Korrelationskoeffizienten befinden sich in Abb. 20. Davon liegen alle Werte (außer der Diagonale) in der Nähe von Null, deswegen weißt das Eingangsfeature kein Zusammenhang sich miteinander auf. D.h. bei solchem Datensatz importiert man keine objektive Präferenz, die eine wichtige Voraussetzung beim Trainieren des Vorhersagemodells ist. Trotzdem bei manchen Situationen würden einige richtige Vorwissen zum Datensatz geführt (Siehe Kapitel 4.3).



Abb. 19: Scatter-Matrix des Eingangsfeatures



Abb. 20: Korrelationskoeffizienten des Eingangsfeatures

## 4.2 Statistische Verteilung

Die Kerndichteschätzung für jedes einzelne Feature ergibt sich in Diagonale in Abb. 19, die ist ein statistisches Verfahren zur Schätzung der Wahrscheinlichkeitsverteilung einer Zufallsvariable. Im Histogramm wird die Häufigkeit einer Zufallsvariable innerhalb einem bestimmten Abstand dargestellt. Deswegen sind die Werte von Häufigkeiten unstetig, die Anzahl der Häufigkeiten ist vom Abstand abhängig. Im Vergleich dazu weist die Stetigkeit in Kerndichteschätzung auf. Durch die auf der Diagonale liegenden Kerndichteschätzungskennlinien kann man sagen, dass die Verteilungen des Eingangsfeatures angenähert als Gleichverteilung angesehen werden kann. Der Grund dahinter ist es, dass die Kerndichteschätzung am meisten Falle quasi als konstante betrachtet werden kann. Für die Vereinfachung besteht eine Anforderung, dass das auf wissensbasierte Vorhersagemodell gegen Ungleichgewicht robust sein muss (Siehe Kap. 4.3).

Wenn ein System linear wäre, würde es die Überlagerungseigenschaft erfüllen. D.h. wären alle Variable mit Gleichverteilung zufällig verteilt, würden die Funktion, die durch lineare Manipulationen von diesen Variablen erzeugt, auch eine Gleichverteilung aufweisen.

Wie in Kap. 3.1 gezeigt ist eine FEM-Analyse ein System von partiellen Differentialgleichungen, sondern nicht als lineare System berücksichtigt werden kann. Im Hinblick auf Verteilung besitzen die Zielgröße i.d.R. andere Profile als die von Eingangsfeature. Zusammenfassend führt die Gleichverteilung des Eingangsfeatures nicht zu einer Gleichverteilung einer FEM-Analyse. Die entsprechende Auswirkung wird in Abb. 21 gezeigt.

Links oben und rechts unteren steht die Kerndichteschätzung, jeweils für maximale Spannung und maximale Verschiebung. Die Beide verfügen sich über einem Ungleichgewicht, insbesondre bei max. Verschiebung. Die meisten Daten von max. Verschiebungen liegen innerhalb einem Bereich von 0 bis etwa 15mm. Trotzdem befindet die Maximale sich über 250mm. Das Ungleichgewicht wirkt sich maßgeblich auf die Genauigkeit des Vorhersagemodells. Darüber wird es in Kap. 4.4 untersucht.

Die restliche zwei Diagramme zeigen die Korrelation zwischen die zwei Zielgrößen. Davon kann es direkt ausgelesen, dass die max. Spannung und max. Verschiebung einen positiven linearen Zusammenhang sich miteinander haben, weil die Datenpunkte auf eine mit positive Steigerung kenngezeichnete schräge Linie zusammenziehen.

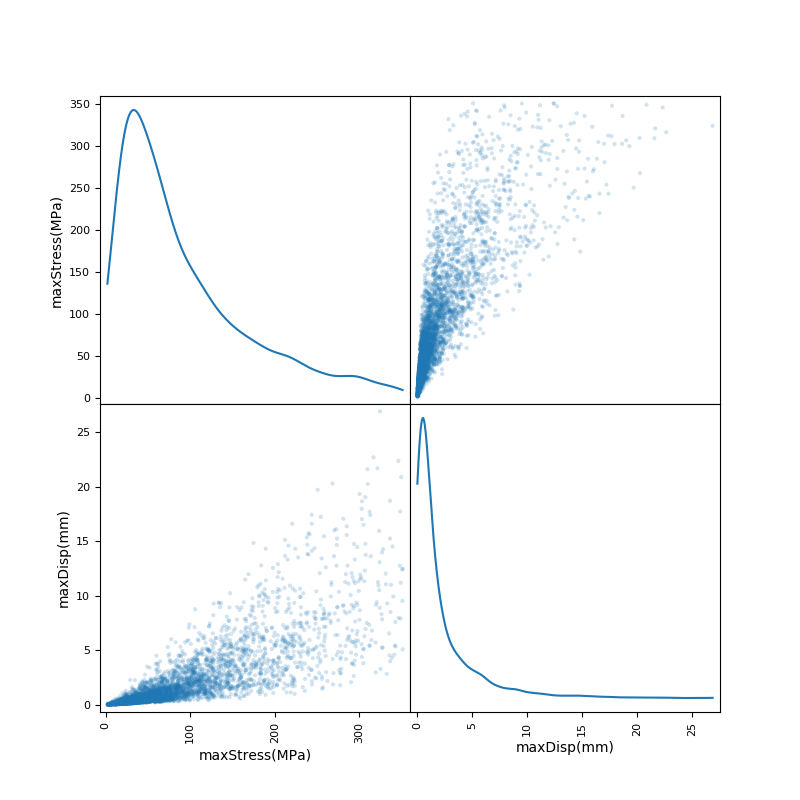


Abb. 21: Scatter-Matrix von Zielgrößen

Zusammenfassend könnte das Vorhersagemodell, das durch einen unausgeglichenen Datensatz trainiert wäre, auch das Ungleichgewicht vom Datensatz erlernen. Z.B. bei dem Datensatz in Abb. 21 konzentrieren die max. Verschiebungen auf kleiner als quasi 15mm. Dazu entsprechende Vorhersagemodell neigt auch zum Vorhersagen kleiner als 15mm. Wegen des Ungleichgewichts des Datensatz ist das Vorhersagemodell nicht in der Lage, die Minderheit mit hoher Genauigkeit vorherzusagen.

In Abb. 22 ist ein Verlauf der Modellgüte des Vorhersagemodells, die durch ausgeglichen Datensatz trainiert ist. Aus diesem Diagramm kann die Ungenauigkeit bei Minderheit (max. Verschiebungen großer als etwa 15mm) deutlich herausgelesen werden.



Abb. 22: Modellgüte bei einem unausgeglichenen Datensatz.

## 4.3 Verstärkung bzw. Ausgleich des Datensatz

Wie in Kap. 4.1 gesagt ein richtiger Datensatz sollte keine objektive Präferenz hereinlassen. Trotzdem zählt Ausgleich des Datensatz nicht dazu, d.h. es ist wissenschaftlich subjektive Verstärkung.

„Studies have shown that for several base classifiers, a balanced data set provides improved overall classification performance compared to an imbalanced data set” [19]

In einer Arbeit von H. He „Learning from Imbalanced Data“ wurden Drei Methoden gegen Ungeleichgewicht vorgeschlagen, nämlich „Sampling Methods for Imbalanced Learning“, „Cost-Senstive Methods for Imbalanced Learning“ und „Kernel-Based Methods and Active Learning Methods for Imbalanced Learning“. Davon sind Oversampling und Undersampling in der künstlichen Intelligenz weitverbreitete. Zusammenfassend bei Oversampling werden einige zufälligen Daten von der Minderheit dupliziert, im Gegensatz dazu bei Undersampling werden einige zufälligen Daten von der Mehrheit entfernt. Deswegen haben die zwei Methoden eindeutige Nachteile. Bei Oversampling entsteht eine Overfitting-Gefahr (siehe Abb. 8) und Undersampling führt zur Gefahr, manche wichtigen Informationen sich verloren zu lassen.

Eine Voraussetzung hinter der obengenannten Methode ist es, dass das Datensatz fixiert ist. D.h. das Wissen für Trainieren eines Modells ist vorgegeben, das könnte schwer oder unmöglich noch mehr angesammelt werden. Dieses Problem bei der Minderheit ist noch schlimmer, deswegen ist weitere Ansammlung der Minderheit im Praxis normalerweise sehr aufwändig. Ein typisches Beispiel ist „Mammography Data Set“, eine Sammlung von Bildern aus einer Reihe von Mammographie-Untersuchungen, die an verschiedenen Patienten durchgeführt wurden. Dieses Datensatz beinhaltet 10923 „Negative“ (gesund) und 260 „Positive“ (krebsartig), dazu lässt sich die Minderheit „Positive“ im Praxis quasi unmöglich gleich wie „Negative“ anwachsen.

Die Drei von H. He vorgeschlagenen Methoden könnten so zusammengefasst, das vorgegebene fixierte Datensatz möglichst vernünftig einzusetzen oder durch Algorithmus zu ergänzen.

Bei uns in der Arbeit ist die Situation anders, d.h. im Gegensatz dazu ist unser FEM-Datensatz flexibel, sondern nicht fixiert und vorgegeben. Deswegen kommt eine neue aber einfache und direkte Lösung gegen Ungleichgewicht, die Minderheit mehr zu generieren.

● Addieren einen zufälligen kleinen Wert zu Eingangsfeature der Minderheit

Aus dem unausgeglichenen Datensatz werden die Datenpunkte der Minderheit angefasst. Die direkt vom Datensatz ausgelesenen Datenpunkte sind als Originale in Abb. 23 mit roten Punkten angesehen. Danach wird eine Zufallsvariante zu deren Eingangsfeature addiert. Die dazu addierte Zufallsvariante besitzt eine Standardnormalverteilung, deren Dichtefunktion sich ergibt durch:

|  |  |
| --- | --- |
|  | (32) |

Die neue erzeugte Eingangsfeature wird durch folgende Gleichung gegeben und in Abb. 23 mit blauen Dreiecken dargestellt:

|  |  |
| --- | --- |
|  | (33) |



Abb. 23: Ausgleich des Datensatz durch Addieren

Die neue Eingangsfeature liegt in der Nähe von Originale und wird in der FEM-Analyse durch automatisiertes Verfahren durchgeführt. Zum Schluss kommen die dazugehörte FEM-Ergebnisse (max. Spannung und max. Verschiebung) heraus. Das Verbinden von Eingangsfeature und die zwei FEM-Ergebnisse bzw. Zielgrößen wird als ein ergänzter Datensatz angesehen und im originalen Datensatz hinzugefügt. Damit wächst die Minderheit an und glicht der originale Datensatz aus.

Würde die Minderheit in einem bestimmten Raum sich befinden, könnte das Addieren einer Zufallsvariante dazu das Raum erweitern.

● Interpolation in Eingangsfeature der Minderheit

Interpolation ist ein Begriff in der numerischen Mathematik. Zu gegebenen diskreten Daten soll eine stetige Funktion gefunden werden, die diese Daten abbildet. Man sagt dann, die Funktion interpoliert die Daten.

Im Gegensatz zum Addieren einer Zufallsvariante begrenzt die sich bei der Interpolation auf das bestimmte Raum, dessen Grenz durch die Minderheit ausgebildet ist. Wie in Abb. 23 und 25 gezeigt steht die Originale (Minderheit) punktweise diskret im Raum zur Verfügung. In der Arbeit entspricht die Interpolation nicht 100% der mathematischen Definition. D.h. uns interessieren die Zielgrößen (max. Spannung und Verschiebung) in diesem Schritt nicht. Hauptsache ist es die neuen Datenpunkte von Eingangsfeature zu erzeugen. Danach gehen die neuen erzeugten Datenpunkte von Eingangsfeature in der automatisierten FEM-Analyse herein und kommen die entsprechenden FEM-Ergebnisse bzw. Zielgrößen (max. Spannung und Verschiebung) aus. Damit wird das Anwachsen der Minderheit realisiert.

Die konkrete Vorgehensweise läuft so (Abb. 24):

1) Für jeden Datenpunkt von Eingangsfeature (roter Punkt in Abb. 25)werden andere zufälligen Punkte bestimmen.

2) Für jede Verbindung von und wird der Mittelpunk berechnet. Es ergibt sich durch:

|  |  |
| --- | --- |
|  | (34) |

Neben dem Mittelpunkt könnte auch Variante z.B. ein Dritte ausgewählt werden.

In Abb.25 werden die interpolierten Datenpunkte von Eingangsfeature dargestellt (bei n=2). Davon sind die grünen Rechtecke die erzeugten Datenpunkte. Die dazu entsprechende Lage ist gleich wie erwartet, d.h. innerhalb des bestimmten Raum von Minderheit.

Durch Addieren eine Zufallsvariante würde das Raum von Minderheit erweitert. Im Gegensatz dazu würde es durch Interpolation das Raum von Minderheit verfeinert.

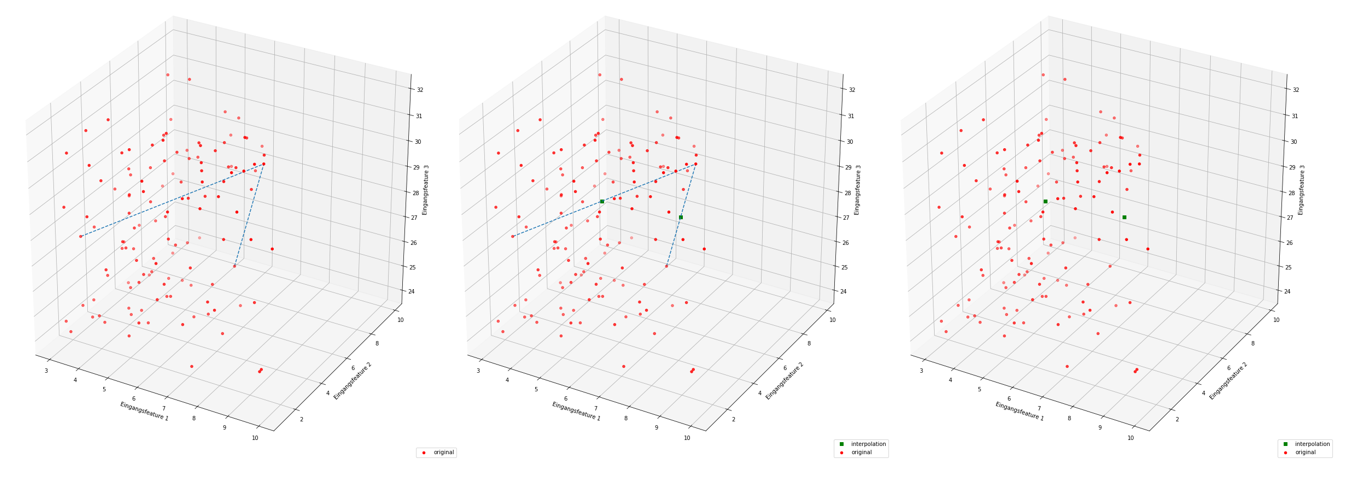


Abb. 24: Vorgehensweise der Interpolation



Abb. 25: Interpolation von Eingangsfeature der Minderheit

## 4.4 Auswirkung der Verstärkung des Datensatz

In Abb. 26 ist die statistische Verteilung mit Verstärkung des Datensatz gezeigt. Solches Diagramm wurde schon in Kap. 4.3 erläutert. Im Vergleich zu Abb. 20 in Kap. 4.1 hat die Kerndichteschätzung von max. Verschiebung eine Verbesserung des Gleichgewichts, weil zwei Bogen im Verlauf der Kerndichtschätzung (rechts unteren) auslesbar sein können.

Trotzdem besitzt das Datensatz noch kein optimales Gleichgewicht, d.h. nach der Verstärkung zählen die Datenpunkte, die mit großen Werte versehen, auch zu Minderheit. Der Grund liegt daran, dass obwohl das Datensatz nicht 100% optimal verstärkt wird, damit kann auch eine deutliche Verbesserung der Genauigkeit bzw. Modellgüte des Vorhersagemodells geschafft werden. Darüber wird es im Folgendes bespricht.

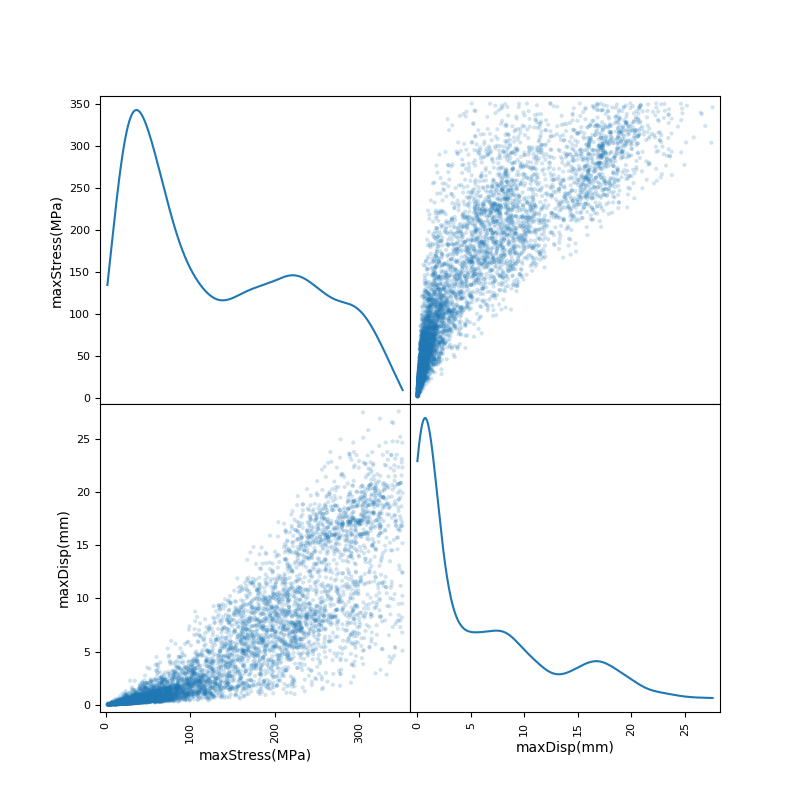


Abb. 26: Verteilung mit Verstärkung des Datensatz

In Abb. 26 werden zwei Diagramme für Modellgüte dargestellt, nämlich links mit Verstärkung und rechts ohne Verstärkung des Datensatz. Die X-Achse steht für die realen Werte, die direkt aus der automatisierten FEM-Analyse. Die Y-Achse steht für die Vorhersage, die direkt aus dem Vorhersagmodell.

Die Abweichung zwischen realer Wert und Vorhersage könnte mit vielseitigen Kennwerten kenngezeichnet werden (siehe Kap. 2.1.1), MAPE und MAE sind zwei davon. MAPE lieget im Titel des Diagramms und MAE ist durch die zwei grünen Linien kenngezeichnet.

Im rechts ist ein Trend erkennbar, dass mit steigenden realen Werte die Abweichung dazwischen sich vergrößert. Der Trend passt zum Aussage von Kap. 4.2. D.h. bei Minderheit und Mehrheit entsteht eine inhomogene Abweichung. Im Gegensatz dazu bei der Verstärkung des Datensatz trifft eine homogene Modellgüte im ganzen Wertebereich auf. Die Kennwerte MAPE und MAE vom links sind deutlich besser als die vom rechts.

Zusammenfassend kann man sagen, dass die Verstärkung bzw. Ausgleich des Datensatz eine Rolle als Booster des Modells spielt.

Theoretisch und mathematisch ist die positive Auswirkung der Verstärkung auf Modellgüte schwer zu beweisen und begründen. In der Arbeit wurde nur eine praktische Validierung geschafft.

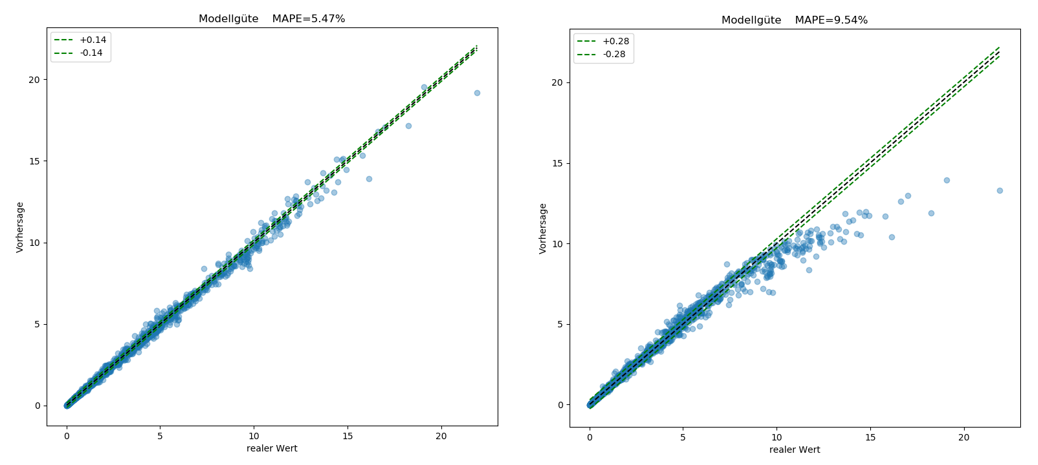


Abb. 26 Vergleich zwischen mit und ohne Verstärkung des Datensatz bei demselben Vorhersagemodell. (links) Modellgüte mit Verstärkung. (rechts) Modellgüte ohne Verstärkung.

# 5 Daten vorbereiten

## 5.1 Datenbereinigung und Ausreißererkennung

Im Praxis enthält das Datensatz normalerweise viele Werte, die das lernfähige Algorithmus nicht erkennen kann. Z.B. NULL-Wert und Text-Wert. Dafür müssen derartige Werte beim Datenvorbereiten bereinigt werden.

Würde ein Eingangsfeature lediglich aus einen festen Wert bestehen, hätte dieses Eingangsfeature keine Auswirkungen auf Trainieren des Vorhersagmodells, d.h. kein Positive und kein Negative, sondern nur relativ mehrere Rechenzeiten. Dazu entspricht in der Arbeit beispielhaft die Elementsgröße der Vernetzung bei FEM-Analyse. Die ist auf 0,2mm festgestellt und daher im Datensatz zum Trainieren entfernt.

Die Werte eines Eingangsfeatures, die deutlich größer als die meisten sind, zählen zum Ausreißer. Die sind i.d.R. durch ein falsches Verfahren und/oder defekte Sensoren ermittelt. Da müssen die beim Datenvorbereiten herausgefunden und entfernt werden.

Die fehlende Werte (NULL und NaN) werden häufig durch falsch Kommunikation während der Datenerfassung erzeugt. Die sind schwer erneut genau zu bestimmen. Die direkte Lösung liegt daran, dass die die fehlende Werte enthaltende Daten ganz entfernt werden können. Die explizite Methode ist mit Mittelwert die fehlende Werte zu ersetzen. Jonathan Sterne wurde eine Methode für die fehlenden Werten in 2009 aufstellen, die heißt Multiple Imputation [20].

In Kap. 3.1 wird das Material in der Arbeit auf „AISI 1020“ festgelegt. Falls mehrere Varianten des Materials untersucht werden wollen, dann geht das Eingangsfeature „Material“ im Datensatz herein (siehe Tabelle 4 links). Trotzdem versteht das Vorhersagemodell die zwei Wörter „AISI 1020“ und „AISI 304“ nicht. Es kann nur mit Zahlwerte arbeiten. Eine Lösung dazu ist für jedes Material ein Eingangsfeature zu erzeugen (siehe Tabelle 4 rechts). Die Zahlwerte 1 und 0 werden zur Charakterisierung des Materials eingesetzt. Wäre ein großes Projekt mit hunderten Materialien, würde diese Lösung zur Dimensionskatastrophe führen.

Tabelle 4: Eine Lösung zur Text-Behandlung





Ausreißererkennung funktioniert als ein Filter, mit dem die Ausreißer herausgefiltert und die Normale bleibt werden können. Mittels des Box-Diagramms in Abb. 27 sind Ausreißer schematisch dargestellt. Die Datenpunkte, die entfernt von die meisten liegen, können als Ausreißer betrachtet werden. Damit kann eine feste Grenz bestimmt werden. Innerhalb dieser Grenz stehen alle normale Datenpunkte. Die Grenz ist als so genannte Filter angesehen. Beispielsweise in Abb. 27 wird die Grenz auf 150 festgelegt. Dann die Werte, die größer als 150 sind, gehören zum Ausreißer und werden vom Datensatz entfernt.

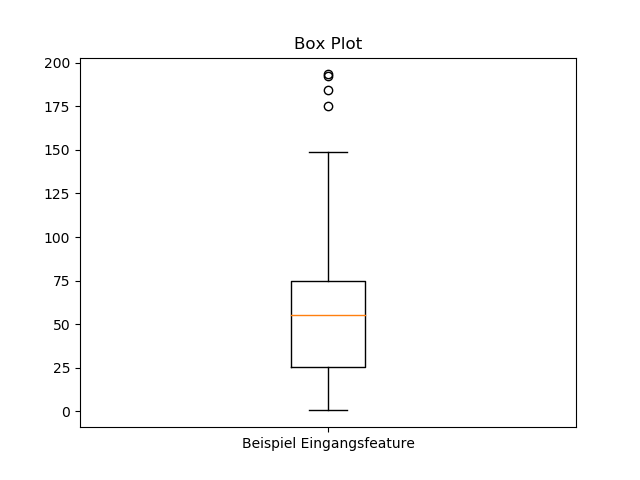


Abb. 27: Boxplot zum Erkennen des Ausreißers

## 5.2 Feature Scaling

I.d.R. verfügen Eingangsfeature unterschiedlichen Maßstab. Z.B. in Tabelle 3 geht das Eingangsfeature „thickness“ von 1 bis 10, aber das Eingangsfeature „F1“ von 0 bis 50. Die Empfindlichkeit der Kostfunktion (Siehe Kap. xxx) gegen Änderung des Gewichts unterschiedet sich wegen unterschiedliches Maßstabs.

In Abb. 28 wird die Auswirkung von Feature Scaling auf Gradientenabstiegsverfahren (siehe Kap. 2.1.2) dargestellt, nämlich links mit Feature Scaling und rechts ohne Feature Scaling. Von Rot bis Grün nimmt die Kostenfunktion ab. Weil Eingangsfeature „thickness“ kleiner ist, braucht es eine größere Änderung von , die Kostenfunktion zu verändern. Deswegen ist die Hauptachse der Ellipse entlang der x-Achse, darauf Eingangsfeature „thickness“ steht.

Im Links geht das Gradientenabstiegsverfahren direkt nach dem Minimum der Kosten, deswegen kommt es schneller in Minimum an. Obwohl kann das Gradientenabstiegsverfahren auch in Minimum ankommen, geht es fest senkrecht zum Minimum am Anfang. Solche Laufbahn verursacht Zeitaufwand und mehre Rechenleistung.

Ziel von Feature Scaling ist es, dass den Maßstab von allen Eingangsfeature zu vereinigen. Es gibt Sonderfälle, z.B. für Entscheidungsbaum ist Feature Scaling nicht benötigt.

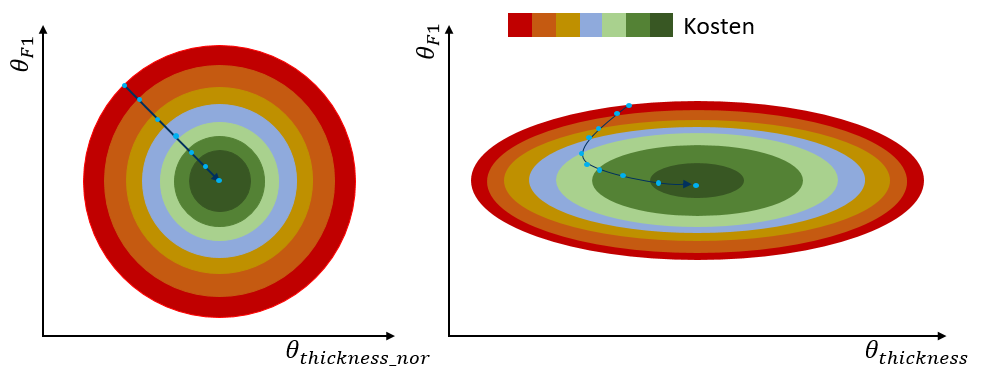


Abb. 28: Gradientenabstiegsverfahren

mit (links) und ohne (rechts) Feature Scaling

● Normalisierung (Min-Max-Scaler)

Alle numerischen Werte des Eingangsfeatures werden auf dem Bereich von 0 bis 1 projiziert. Abb. 29 zeigt dieses Verfahren. Die mathematische Formel ergibt sich durch:

|  |  |
| --- | --- |
|  | (35) |

Da *m* ist die Anzahl der Trainingsdaten.

Damit können einheitsspezifische Werte zu reinen numerischen Zahlenwerte umwandeln. Weiterhin sind die vom Maßstab unabhängig.

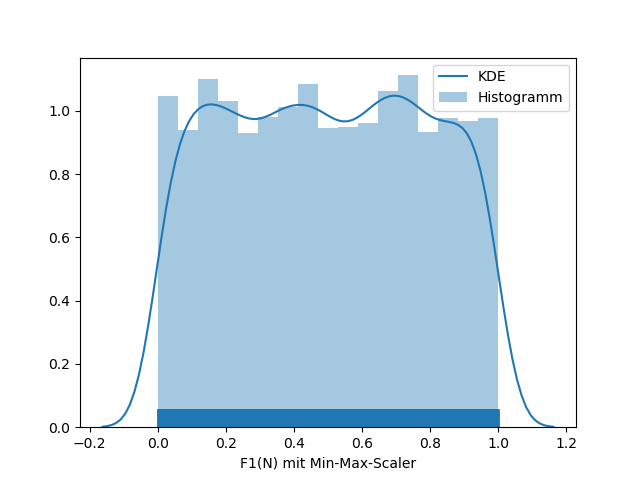


Abb. 29: Verfahren der Normalisierung,

(Links) Originale im Wertebereich von 0 bis 50,

(Rechts)Mit Min-Max-Scaler im Wertebereich von 0 bis 1.

● Standardisierung

Im Vergleich zur Normalisierung projiziert es bei der Standardisierung all originalen Werte auf einem bestimmten Bereich, wo in der Nähe von 0 liegt. Die Transformation wird in Abb. 30 dargestellt und ergibt sich durch:

|  |  |
| --- | --- |
|  | (36) |

Da steht für den Mittelwert, steht für die Standardabweichung, *m* ist die Anzahl der Trainingsdaten.

Mit Hilfe dieser Transformation treten die meisten sich in der Nähe von 0 und sich über eine Standardabweichung von 1 verfügbar.

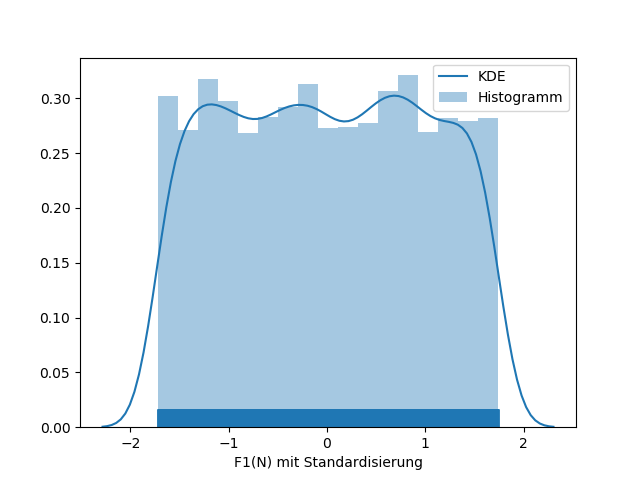


Abb. 30: Verfahren der Standardisierung,

(Links) Originale im Wertebereich von 0 bis 50,

(Rechts)Mit Standardisierung in der Nähe von 0.

Die obere genannte zwei Scaler sind lediglich linear proportionale Manipulation, d.h. die eingesetzten Koeffizienten (Minimum, Maximum, Mittelwert und Standardabweichung) sind für die Scaler entscheidet. Diese Koeffizienten stammen aus den Trainingsdatensatz. Beim Testen und Anwenden des Vorhersagemodells müssen dieselbe Koeffizienten aus Trainingsdatensatz immer eingesetzt werden. D.h. man sollte nicht für Testdatensatz und Anwendungsfälle neue Scaler-Koeffizienten einstellen.

# 6 Vorhersagemodell erstellen

## 6.1 SVM Model für ein Balkenbauteil unter Lasten

Der Aufbau des Trainingsdatensatz ist in Kap. 3.2 gezeigt. In der öffentlichen Bibliothek Scikit-Learn ist eine SVM Funktion bereit verfügbar. Nach benötigten Datenvorbereitungen (siehe Kap. 5) lauft SVM Modell in der Gittersuche iterative durch, um ein mit besten Performance verfügtes Modell zu bestimmen.

Ein vordefiniertes SVM Modell stellt die veröffentlichte Bibliothek Scikte-Learn bereit, nämlich [6]:

*SVR(kernel=’rbf’, degree=3, gamma=’auto\_deprecated’, coef0=0.0, tol=0.001, C=1.0, epsilon=0.1, shrinking=True, cache\_size=200, verbose=False, max\_iter=-1)*

Damit könnte die wichtigsten Hyperparameter, nämlich Kernel, Strafterm C, maximale Iterationen, eingestellt werden. Die ganzen Beschreibungen von anderen Variablen zeigt die Scikit-Learn-Dokument.

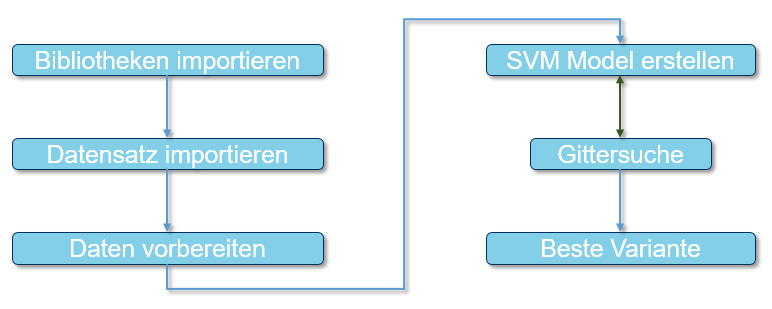


Abb. 31: Laufbahn zum Erstellen eines SVM Modells



Abb. 32: Abschnitt von SVM Code

Abb. 32 stellt einen Teil von SVM Code dar. Basiert auf die Scikit-Learn braucht es eigentlich weniger als 10 Zellen Code, ein SVM Modell zu erstellen. Zuerst lassen die benötigten Bibliotheken herein importieren. Dann wird Datensatz importiert und für SVM vorbereitet. Nach Erstellen eines leeren Modell läuft die Gittersuche mehrmals jeweils mit unterschiedlichen Hyperparameter, um die beste Variante zu bestimmen. Hier ist die MAE als die Bewertung der Modellgüte genommen.

Um die besten Hyperparameter zu bestimmen netzt man i.d.R. eine Gittersuche, wobei werden alle Kombinationen der vorgegebenen Hyperparameter nacheinander erledigt. Danach könnte man die besten Parameter, die über beste Modellgüte verfügen, bestimmen.

Die Modellgüte noch verbessern zu können wird eine feine Gittersuche in der Nähe von der Beste durchgeführt.

Abb. 33 zeigt das Ergebnis bzw. die Modellgüte eines SVM-Vorhersagemodells für einen Balken unter vertikalen Lasten. Davon erreicht MAPE auf 5%, d.h. die Vorhersage weicht sich von den realen Werten um 5%. In anderen Worten auf Basis von vorgegeben Wissen kann das SVM Modell bei solchen Situationen mit ausreichender Genauigkeit die FEM-Zielgröße (max. Verschiebungen) vorhersagen.

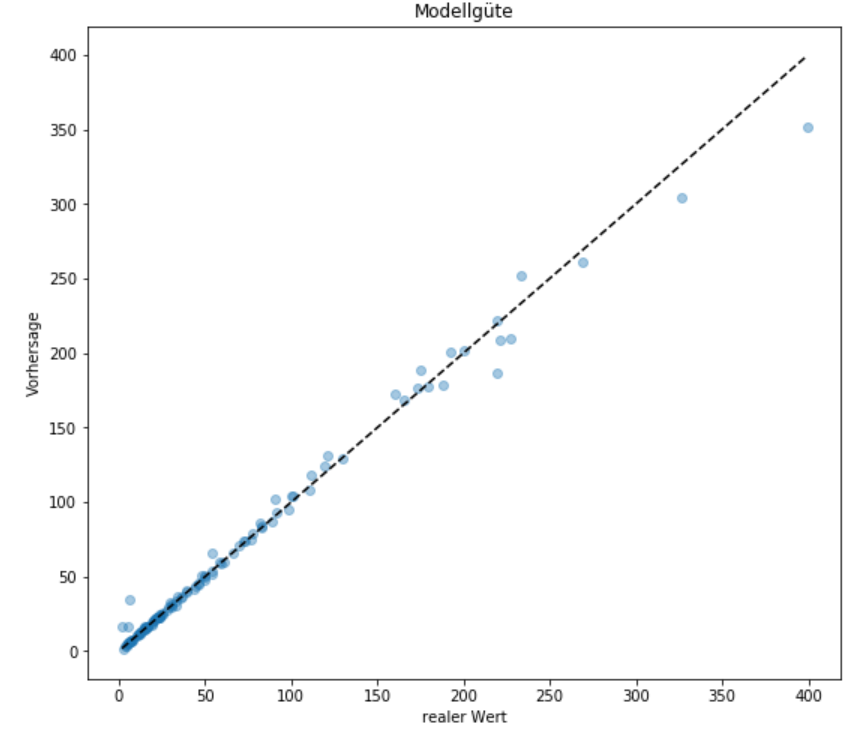


Abb. 33: SVM Modellgüte, MAPE=5%

## 6.2 Neuronales Netz für W-Profil-Bauteile unter Lasten

Solche Bauteile zeigt Abb. 1. Den Aufbau des entsprechenden Datensatz erläutert es in Kap. 3.2. In der Arbeit wird die veröffentlichte Bibliothek TensorFlow Keras als Werkzeug, das neuronale Netz zu erstellen [4].

Zunächst wird einige grundlegende Elemente in Keras vorgestellt, welche beim Erstellen des Modells zum Einsatz kommen [5].

*Sequential():*

Die Elemente werden in eine lineare Reihenfolge nebeneinander angeordnet, wobei geht ein Beispiel von Eingangsschicht zur Ausgangsschicht durch.

*keras.layers.Dense(units, activation=None, use\_bias=True, kernel\_initializer='glorot\_uniform', bias\_initializer='zeros', kernel\_regularizer=None, bias\_regularizer=None, activity\_regularizer=None, kernel\_constraint=None, bias\_constraint=None):*

auf Engl. densely/fully connected. Das ist eine neuronale Ebene, worum wird jedes Neuron o. Knoten vollständig verbunden. Die Beschreibung ist Gegensatz zum nicht vollständigen Verbunden, z.B. Convolutional Layer im Convolutional Neural Network (CNN).

*kernel\_initializer:*

womit werden die initialen Gewichte der Verbindungen bestimmt. Die können zufällig eingestellt werden, aber mit ein verstärktes Verfahren kann eine schnelle Konvergenz erreicht. In der Arbeit wird *keras.initializers.glorot\_uniform*  genutzt. Mehr Details dafür stehen in der Literatur [21].



Abb. 34: Abschnitt von NN Code

Abb. 34 zeigt einen Abschnitt von NN Code. Die Laufbahn ist ebenso gleich wie die von Erstellen des SVM Modells. Zuerst muss man die benötigten Bibliotheken herein importieren, dann das Datensatz zu importieren und für das NN Model zu vorbereiten. Mit der Funktion *create\_model* kann NN Model durch Hyperparameter günstig erstellt werden. In der Gittersuche werden mehrmals die Varianten von Kombination der Hyperparameter trainiert. Hier wird der MAPE Kennwert als Kostenfunktion genommen, der wird auch als Kriterium zur Charakterisierung der Modellgüte. Zum Schluss wird die Beste davon genommen, zum Vorhersagemodell zu dienen.

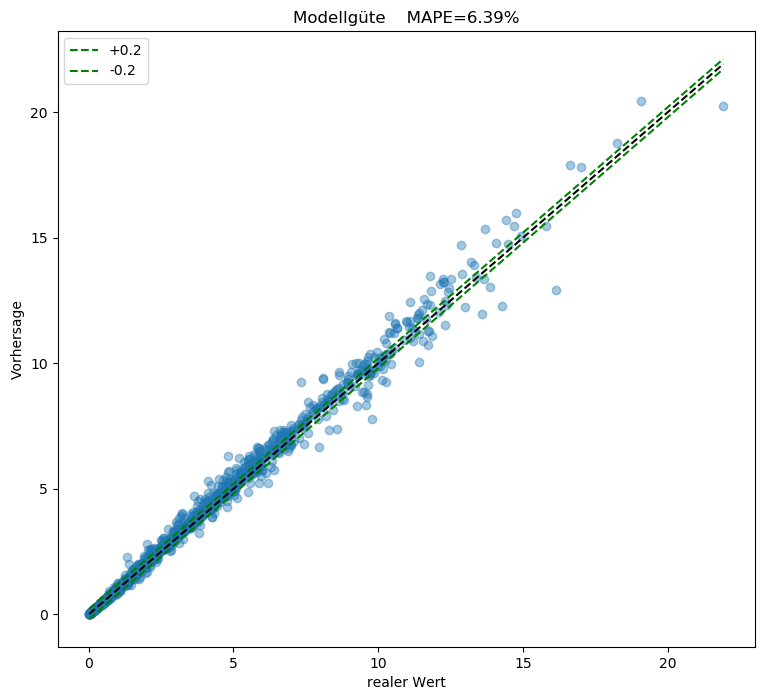


Abb. 35: Modellgüte von NN Modell, MAPE = 6,39%

Nach der Gittersuche in einem bestimmten Bereich tritt das beste Vorhersagemodell auf. Die Modellgüte zeigt Abb. 35. Die entsprechende MAPE beiträgt 6,39%. Obwohl die Genauigkeit schon gut ist, bleibt noch eine schlechte Tendenz im Modell, dass mit dem steigenden Beitrag weist eine sich vergrößernde Abweichung auf. In anderen Worten, mit demselben Modell weist es unterschiedliche Performance auf, welche abhängig vom Wertbereich ist.

## 6.3 Stufiges Modell für W-Profil-Bauteile unter Lasten

Grundprinzip für solche obergenannte Tendenz zu vermeiden ist für jeden Wertbereich sein eigenes Vorhersagemodell zu erstellen. Die Anzahl des Wertbereichs hängt von der Komplexität und Rechenleistung ab. Um die unbekannten Zielgrößen bei der Vorhersage zum richtigen Wertbereich zu zuordnen nutzt man ein Klassifikationsmodell vor dem Regressionsmodell. Den Aufbau des stufigen Modells zeigt Abb. 36. Quellecode wird im Anhang beigefügt.

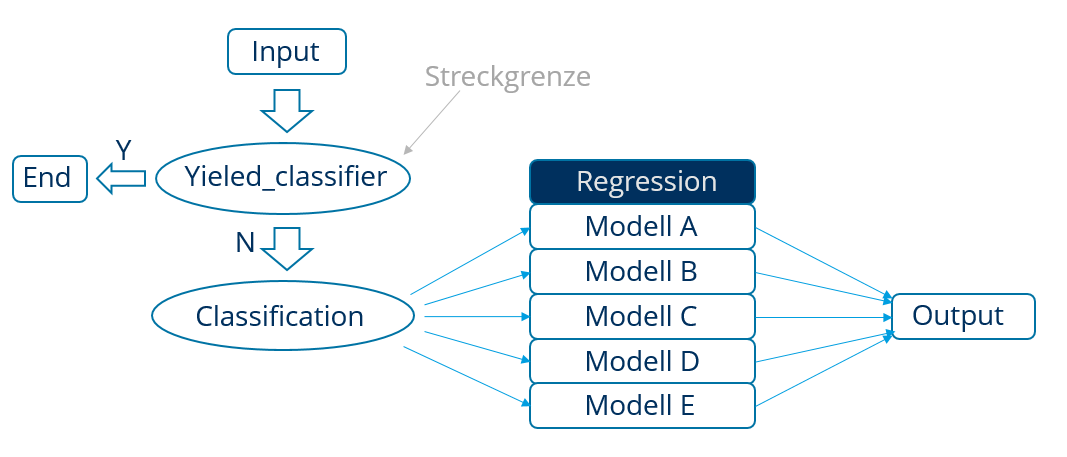


Abb. 36: stufiges Modell

Der Yieled\_classifier macht die Aufgabe, anhand der maximalen Spannung im Datensatz die Streckgrenze zu überprüfen, ob die Bauteile die Streckgrenze überschreiten. Damit wird die Zielgröße (max. Verschiebung) in einem gewissen praxisnahen Wertbereich beschränkt. Aber wegen der Spannungsspritze in der FEM-Analyse macht der Klassifikator nur eine grobe Einschätzung.

Nach dem ersten Klassifikator folgt noch eine Klassifikation, wobei wird es zugeordnet, in welches Regressionsmodell hereinzugehen. In der Arbeit wird es in 5 Regressionsmodelle untergeteilt, nämlich Modell A bis E, deren Wertbereich in Abb. 37 dargestellt.

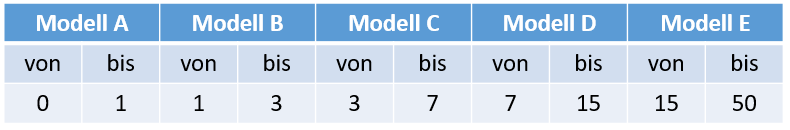


Abb. 37: Wertbereich von Modellen

Für die Bestimmung der Ober- und Untergrenze von Wertbereich liegt es keine expliziten Kriterien zur Verfügung. Hätte es mehr Unterteilen, würde mehre Arbeit benötigt, das entsprechende Regressionsmodell zu erstellen.

Durch Gittersuche beim Trainieren ergeben sich die beste Variante der Modelle, sowohl Streckgrenze-Klassifikator, Klassen-Klassifikator als auch die Regressionsmodelle.

Abb. 38 zeigt das Ergebnis bzw. Konfusionsmatrix von Streckgrenze-Klassifikator, davon beträgt die Genauigkeit 96.5%. Deswegen ist es bestimmt, dass der Streckgrenze-Klassifikator schon erlernt, die Überschreitung der Streckgrenze eines Bauteils zu erkennen.

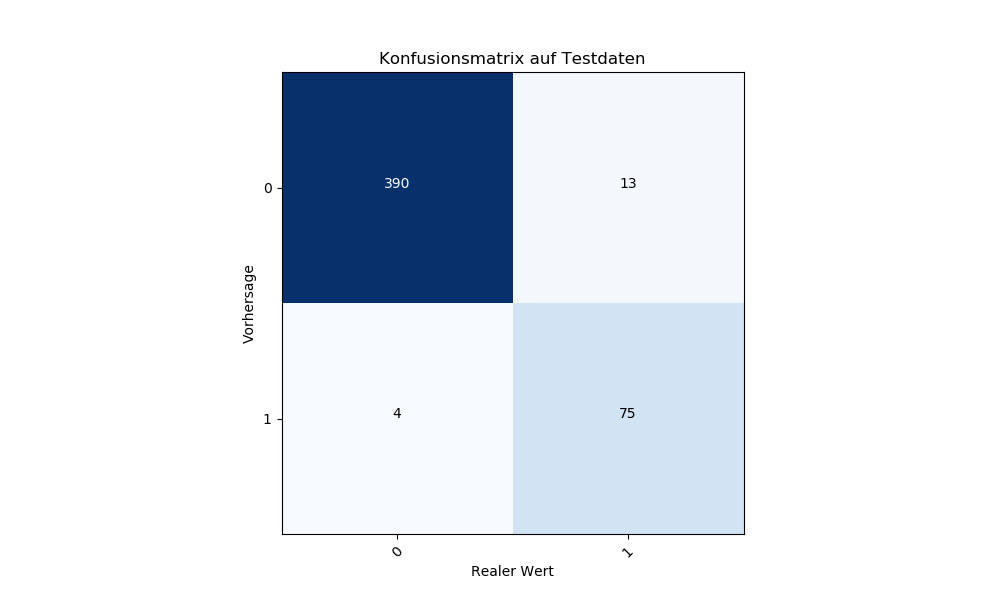


Abb. 38: Konfusionsmatrix von Streckgrenze-Klassifikator

Abb. 39 zeigt das Ergebnis von Klassen-Klassifikator. Die Genauigkeit erreicht 90,6%. Zum einen ist die ausreichend für das nachfolgende Regressionsmodell, zum anderen weißt es noch eine Potenzial auf, diese Genauigkeit noch zu erhöhen.

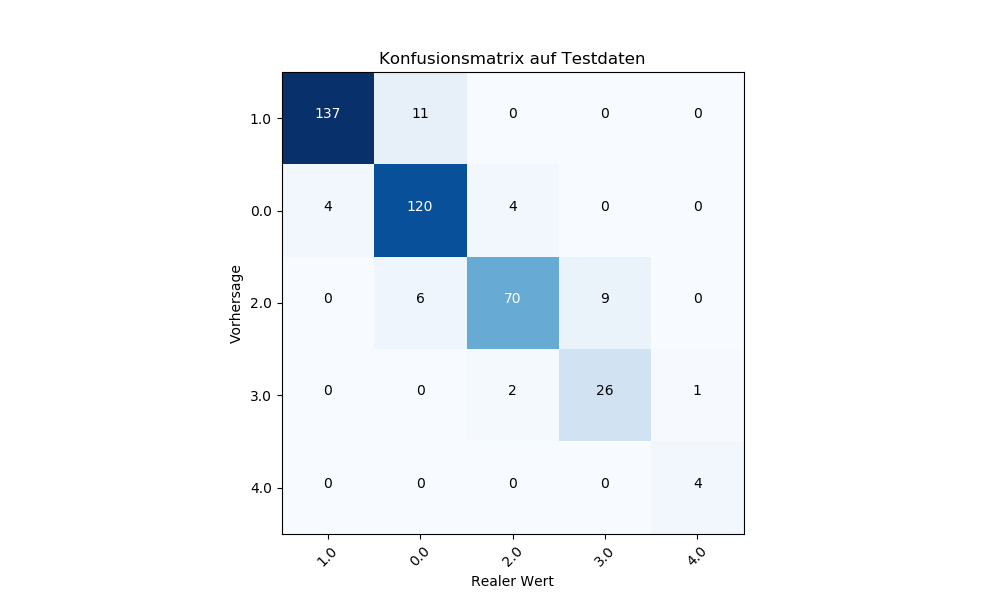


Abb. 39: Konfusionsmatrix von Klassen-Klassifikator

Die letzte Stufe ist das Regressionsmodell, deren Ergebnisse zeigt Abb. 40. Die erreichbare MAPE ist weniger als 5%. Im Vergleich zu einem einzelnen neuronalen Netz ist die Performance des stufigen Modells gleichmäßig, unabhängig vom Wertbereich.

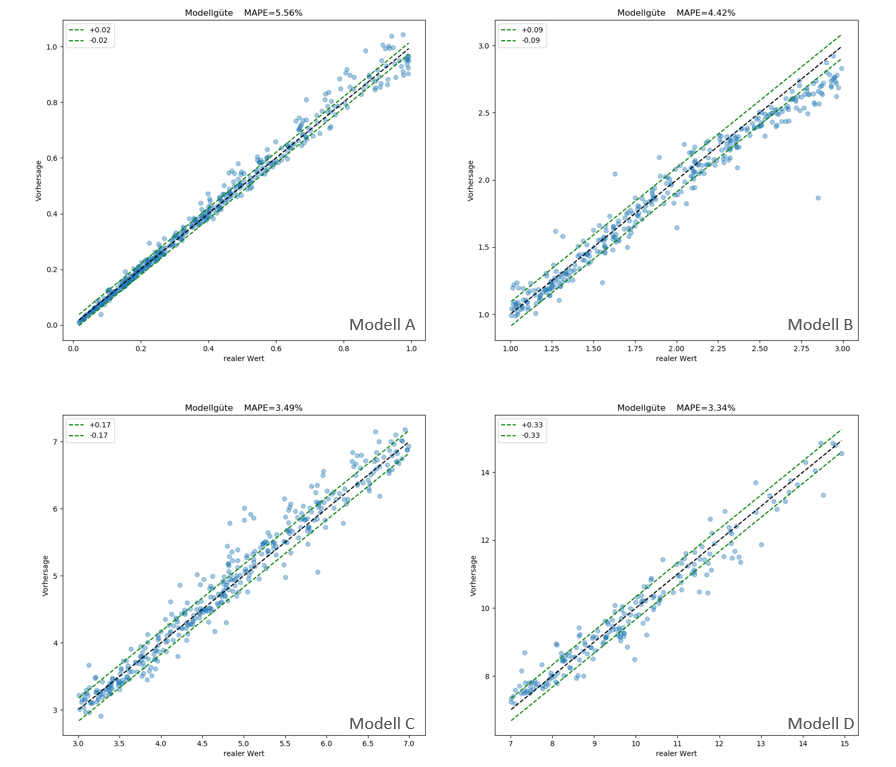


Abb. 40: Regressionsmodelle

Abb. 41 stellt die gesamte Modellgüte auf den ganzen Bereich. Eine Farbe bedeutet ein unterliegendes Regressionsmodell. Die meisten Datenpunkte liegen in der Nähe von der richtigen Linie. Trotzdem gibt es noch Ausreißer. Die Ursache liegt daran, dass die falsche Klassen-Klassifikation zu solchen Ausreißern führt.

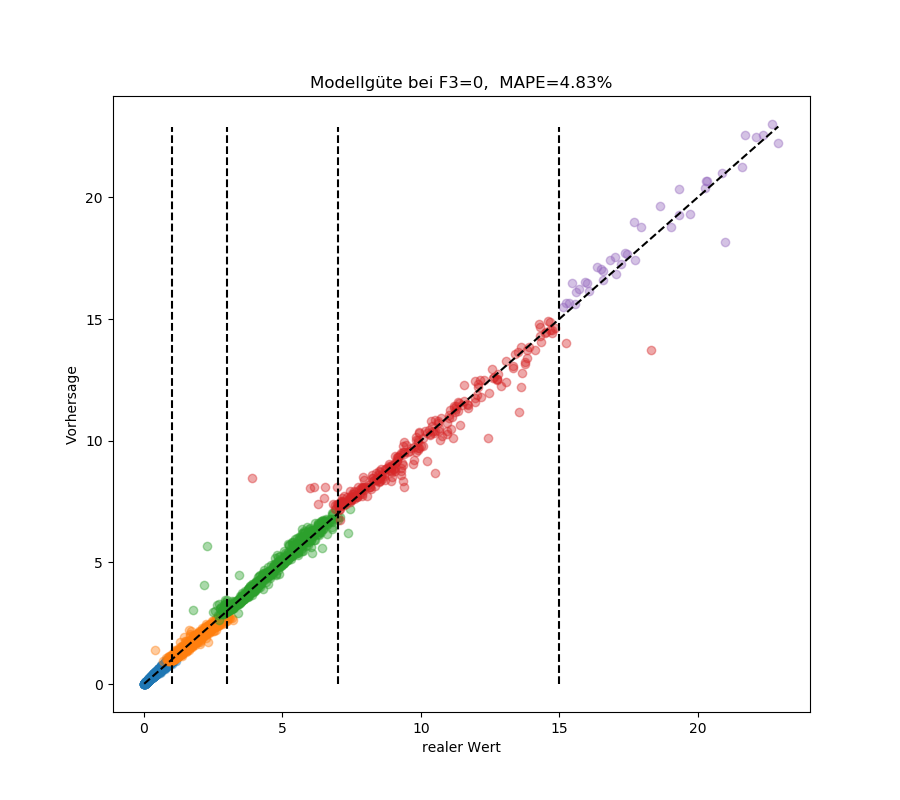


Abb. 41: Modellgüte des stufigen Modells

Zusammenfassend ist das stufige Modell ausreichend in der Lage, die maximale Verschiebung eines W-Profil-Bauteils unter Lasten mit hoher Genauigkeit vorherzusagen. Im Vergleich zu einem einzelnen neuronalen Netz Modell verfügt sich das stufige Modell über eine deutliche Verbesserung.

# 7 Auswertung und Bewertung

Im Vergleich zur FEM-Analyse besitzt das Vorhersagemodell, welche auf Basis von verfügbaren Wissen ist, sowohl Vorteile als auch Nachteile.

Normalerweise konzentriert sich die Rechenleistung während eine FEM-Analyse auf die Vernetzung des Bauteils und das Lösen von Differentialgleichungen. Dagegen macht das auf Wissen basiertes Vorhersagemodell die Vernetzung und das Lösen von DGLn nicht, sondern braucht es mehre Rechenleistung und Zeit beim Trainieren. Bei der Verwendung des Vorhersagemodells, um ein Beispiel vorherzusagen, ergibt sich das Ergebnis maßgeblich schneller als eine FEM-Analyse. In anderen Worten für das Vorhersagemodell ist ein geometrisches 3D oder 2D Modell nicht benötigt, sondern nur die Daten dazu wichtig. Dies entspricht der Aussage, dass das Vorhersagemodell keine Vernetzung und kein Lösen von DGLn macht. Der Einsatz eines Vorhersagemodells erleichtert die Komplexität und Reichenleistung des Algorithmus maßgeblich.

Die Begrenzung des Vorhersagemodells liegt darin, dass i.d.R. nur für solche perimetrischen Bauteile ein bestimmtes Vorhersagemodell erstellen werden könnte. Für die vielseitigen Bauteile ist das Vorhersagemodell nicht geeignet und daran nicht angepasst. Diese Eigenschaft beschränkt die Anwendung auf ein sehr enges Bereich. Neben dem gibt’s auch Schwierigkeit, das Trainingsdatensatz sich anzusammeln. Der Umfang des in der Arbeit verwendeten Datensatz zeigt die Abb. 42. Um tausende Daten zu erfassen, d.h. die FEM-Analyse automatisiert tausendmal durchzuführen, verlangt dies Stunden- bis Tagelang. Falls die geometrischen Profile noch komplexer wären, müssten noch mehr Daten zur Verfügung stellen.

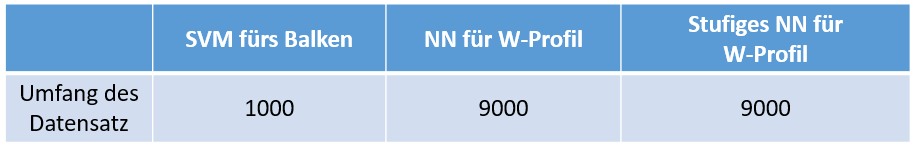


Abb. 42: Umfang des Datensatzes

Unterschied zwischen SVM und NN liegt daran, zum einen basiert sich SVM auf Statistik (Siehe Kap. 2.2.1) und durch Algorithmus wird es zum Konvex-Optimierungsproblem umwandelt. Dazu findet es theoretisch bestimmt das Optimum, zum anderen ist es lediglich für klein Datensatz geeignet, weil behandelt es sich um die Operation von m-Ordnung-Matrix (m ist die Anzahl der Daten). Normalerweise auf Basis von einem bestimmten Datensatz hat SVM eine bessere Verallgemeinerbarkeit als NN. In anderen Worten hat SVM eine klare mathematische Erläuterung.

Aus Beschränkung des Datensatzes für komplexe Geometrie ist NN eigentlich geeignet. Obwohl hat Backpropagation einen großen Fortschritt zum NN mitgebracht, gibt’s noch unklare Funktionsweise darinnen (z.B. man kann noch nicht quantitativ die Konvergenz-Geschwindigkeit analysieren). Deswegen wird NN heutzutage im Praxis als Black-Box angesehen.

Außerdem kann man die fachlichen Kenntnisse oder Erfahrungen ins NN importieren, z.B. für Bilderkennung funktioniert CNN deutlich besser die anderen NN-Strukturen.

In der Arbeit sind das Fachwesen herein importiert, dass die Streckgrenze zur oberen Grenze für die Spannungen dient. Dies Fachwesen (oder Annahme) erleichtert maßgeblich die Anforderung vom Anfang des Datensatzes.

# 8 Zusammenfassung und Ausblick

In der Arbeit am Anfang ist Stand der Technik von maschinellen Lernen untersucht. Dabei werden die mathematischen Grundlagen von SVM und NN erläutert. Anhand des Verfahrens von maschinellen Lernen sollte die benötigten Daten erfasst werden. Durch automatisierte FEM-Analyse bei SolidWorks-Simulation in C# API kommen die Daten an. Die Daten müssten an die Anforderung des Modells angepasst werden, nämlich Datenbereinigung, Ausreißererkennung und Feature Scaling. Zum Schluss macht es eine Gittersuche, die besten Hyperparameter zu bestimmen.

Drei auf Wissen basierten Modelle sind untersucht, nämlich SVM, NN Model, stufiges NN Model. Die Outputs davon annähern sich zur tatsächlichen FEM-Analyse und verlangen sie maßgeblich kürzeren Zeit als FEM-Analyse, vorherzusagen. Im Rahmen der Arbeit ist es bestätigt, dass wissensbasierte Methoden (SVM und NN) ähnlich qualifizierte Aussagen treffen können, wie die numerische Simulation (FEM).

Trotzdem begrenzen sich solche Modelle auf ein sehr enges Anwendungsbereich, d.h. die können die allgemeine FEM-Analyse nicht ersetzen.

In Kap. 3.1 lässt sich das Prinzip einer FEM-Analyse erläutern. In einer FEM-Analyse werden DGLn iterativ gelöst, da konzentriert sich die höchsten Reichenleistungen. Falls ein maschinelles Modell das Lösen der DGL lernen könnte, würde eine FEM-Analyse keine DGLn manipulieren, sondern durch dieses Modell ersetzen und würde die Analyse deutlich beschleunigen.

Bis hier lässt sich die Richtung vom Eingang zum Ausgang eines Modells untersuchen. Eine interessante Frage ist es, was würde ankommen, wenn die Richtung umgekehrt würde. Das würde als ein Vorschlagsystem und Berater funktionieren. Wenn die Ingenieure beispielsweise eine bestimmte Verschiebung eines Bauteils wollten, würde diese Anforderung vom Ausgang zum Eingang durchgehen, die Geometrie und Lasten würden beim Eingang zur Verfügung stehen.

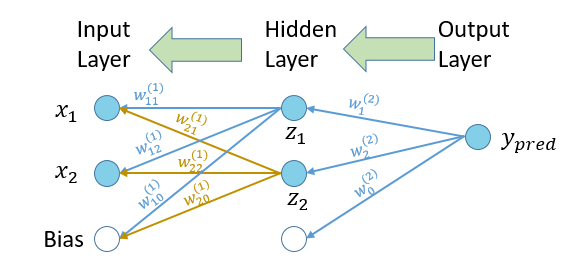


Abb. 43: umgekehrtes NN Model

In den letzten Jahrzehnten hatte künstliche Intelligenz bzw. maschinelles Lernen großen Fortschritt und die Entwicklung dazu lauft noch. Die wirken sich positiv auf Maschinenbaubereich, damit kann man das traditionelle Lösungsverfahren teilweise oder ganz ersetzen, um hoher Lösungsgenauigkeit oder schneller und effizienter Lösungsweg zu erzielen.

# 9 Literaturverzeichnis

[1] Hrennikoff, Alexander. "Solution of problems of elasticity by the framework method". Journal of Applied Mechanics. 8.4: 169–175. 1941

[2] David Kriesel. Ein kleiner Überblick über Neuronale Netze. www.dkriesel.com. 2005

[3] Paul Werbos. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. Dissertation. Harvard University. 1975

[4] Seite „TensorFlow“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 16. Juni 2019, 10:29 UTC. URL: https://de.wikipedia.org/w/index.php?title=TensorFlow&oldid=189583923

[5] Seite „Keras“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 30. Dezember 2018, 11:57 UTC. URL: https://de.wikipedia.org/w/index.php?title=Keras&oldid=184209642

[6] Seite „Scikit-learn“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 20. Februar 2019, 10:36 UTC. URL: https://de.wikipedia.org/w/index.php?title=Scikit-learn&oldid=185863450

[7] Ethem Alpaydin. Introduction to machine learning. 2. Edition. The MIT Press. London, England. 2010

[8] A.K. Jain, M.N. Murty, P.J. Flynn. Data clustering: a review. ACM Computing Surveys (CSUR). Volume 31 Issue 3. 1999.

[9] Aurelien Geron. Hands-On Machine Learning with Scikit-Learn & TensorFlow. First Edition. OREILLY. 2017

[10] C.W. Hsu, C.C. Chang, C.J. Lin. A Practical Guide to Support Vector Classification. Departement of Computer Science. National Taiwan University. Taipei. 2010

[11] P. J. Werbos. „Beyond Regression: New Tools for Prediction and Analysis in

the Behavioral Sciences“. Diss. Harvard University, 1974.

[12] Friedrich U. Mathiak. Die Methode der finiten Elemente (FEM). Neubrandenburger. 2010

[13] Michael Brand. FEM-Praxis mit SolidWorks. Springer Vieweg. 2016

[14] Glenn Whyte. Ins and outs on meshing elements for SOLIDWORKS SIMULATION. https://hawkridgesys.com/blog/ins-and-outs-on-meshing-in-solidworks-simulation. 2018

[15] Seite „C-Sharp“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 24. Juni 2019, 05:26 UTC. URL: https://de.wikipedia.org/w/index.php?title=C-Sharp&oldid=189816714 /16/ A.T. Mathew, C.S.P. Rao. A Novel Method of Using API to Generate Liaison Relationships from an Assembly. J. Software Engineering & Applications. 2010

[17] Seite „Programmierschnittstelle“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 3. Februar 2019, 09:30 UTC. URL: https://de.wikipedia.org/w/index.php?title=Programmierschnittstelle&oldid=185336207

[18] K. Schweizer, E. Boller, G. Braun. Der Einfluß von Klassifikationsverfahren Stichprobengröße und strukturellen Datenmerkmalen auf die Klassifizierbarkeit von Variablen. Methods of Psychological Reasearch Online Vol. 1, No. 4. 1996

[19] H. He, E.A. Garcia. Learning from Imbalanced Data. IEEE transactions on knowledge and data engineering. Vol. 21, No. 9. 2009

[20] Jonathan Sterne. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. https://www.bmj.com/content/338/bmj.b2393. 2009

[21] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In International Conference on Artificial Intelligence and Statistics, 2010.

**Erklärung**

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Werken wörtlich oder sinngemäß übernommenen Gedanken sind unter Angabe der Quellen gekennzeichnet.

Ich versichere, dass ich bisher keine Prüfungsarbeit mit gleichem oder ähnlichen Thema bei einer Prüfungsbehörde oder anderen Hochschule vorgelegt habe.

Ort, Datum Unterschrift