

4. Digital Real Estate Summit 2018

6. März 2018

Workshop: Machine Learning for Commercial Real Estate

Prof. Dr. Manfred Vogel & Lucas Brönnimann
Hochschule für Technik, FHNW



Machine Learning

Es gibt zwei Arten von Machine Learning Algorithmen:

- Supervised Learning:
Es stehen *labeled* (von Menschen bearbeitete) Trainingsdaten zur Verfügung, d.h. es wird versucht, "Expertenwissen" nachzubilden.
- Unsupervised Learning:
Den Algorithmen stehen keine bekannten Zielwerte zur Verfügung. Es wird versucht, Muster in den Daten selbständig zu erkennen.



4 grosse Machine Learning Themen

- Regression:

supervised

 Es werden Regeln und Muster in Daten gesucht, um eine kontinuierliche Grösse vorauszusagen.
 Bsp.: [Versicherungswert von Gebäuden](#)
[Herstellungskosten Maschinenteile](#)
- Klassifikation:

supervised

 Daten werden in zwei oder mehrere im Voraus bekannte Klassen eingeteilt.
 Bsp.: [Gebäude nach Nutzungsart](#)
[Spam und normale E-Mails](#)



4 grosse Machine Learning Themen

- Clustering:

unsupervised

 Daten werden in mehrere Gruppen eingeteilt, welche nicht im Voraus bekannt sind.
 Bsp.: [Gebäude nach Eigenschaften](#)
[Dokumente und Bilder](#)
- Recommender Systeme:

unsupervised

 Nutzer werden auf Objekte (Produkte) aufmerksam gemacht, welche am meisten interessieren dürften.
 Bsp.: [Amazon, Netflix, Spotify, etc.](#)



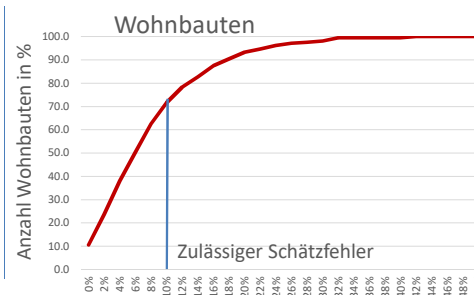
Machine Learning

Lineare Regression:

Geschäftsmodell:

Banken, Versicherungen aber auch Private können Kennwerte (Baukosten, Versicherungswert, etc.) von Immobilien über eine Web-Plattform schätzen.

Je detaillierter die Eingabe, umso genauer die Schätzungen.



Eingesetzte Methoden : Lineare Regression und Kernel-Regression

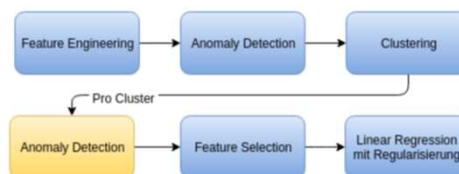


Machine Learning

Lineare Regression:

Einfache ML-Pipeline:

- Eigenschaften der Objekte werden (künstlich) erzeugt.
- Ungewöhnliche Objekte (nicht schätzbare) werden eliminiert.
- Objekte werden in Gruppen eingeteilt (Wohnbauten, Garagen, Schulhäuser, etc.).
- Nochmaliges Testen, ob allenfalls Objekte in falsche Gruppen eingeteilt sind.
- Identifizieren, welche Features wichtig sind resp. welche überflüssig sind.
- Bilden des ML-Modells





Machine Learning

Feature Engineering:

Aus Daten können beliebig viele "Features " generiert werden.

Beispiel:

Für Immobilien seien (neben der Zielgrösse, dem Preis) nur die Wohnfläche F und das Gebäudevolumen V gegeben.

Mögliche generierte Features: $\log(V)$, $F \cdot V$, F^2 , V^2 , F^3 , V^3 , $F^2 \cdot V$, ...

Es können aber raffiniertere Features generiert werden, z. Bsp. der Durchschnittspreis der 15 nächstgelegenen, vergleichbaren Gebäude, die Anzahl der Schwimmings-Pools in der Umgebung (aus Google-Earth), Steuerfuss der Gemeinde, etc.



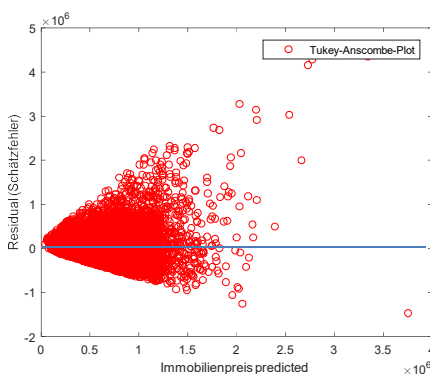
Machine Learning

Beispiel Lineare Regression:

Lineare Regression bedeutet nicht, dass die Schätzgrösse (z. Bsp. der Immobilien-Preis) durch eine Linearkombination gegebener Grössen ausgedrückt wird.

Wie oben gesehen sind beliebige nicht-lineare Kombinationen der Features möglich.

Auch muss die Zielgrösse allenfalls transformiert werden, damit gute Resultate entstehen.



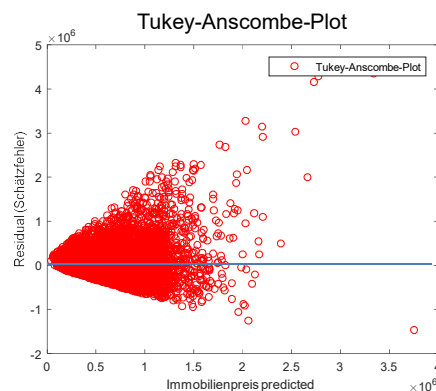


Machine Learning

Beispiel Lineare Regression:

Der nebenstehende Plot zeigt, dass die Schätzungen einen systematischen Fehler haben. Mit zunehmenden Immobilienpreisen nimmt auch der Schätzfehler zu.

Dies kann korrigiert werden, beispielsweise indem nicht der Preis sondern der logarithmierte Preis geschätzt wird.

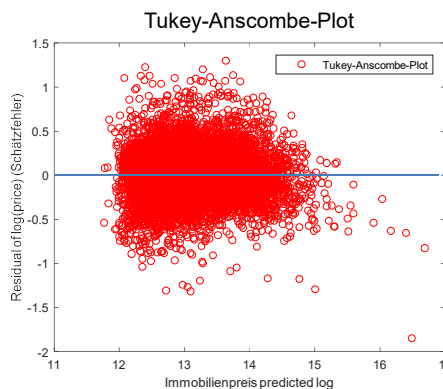


Machine Learning

Beispiel Lineare Regression:

Der nebenstehende Plot zeigt, dass die Schätzungen einen systematischen Fehler haben. Mit zunehmenden Immobilienpreisen nimmt auch der Schätzfehler zu.

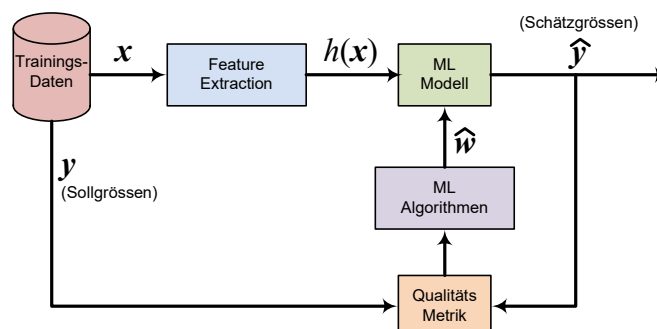
Dies kann korrigiert werden, beispielsweise indem nicht der Preis sondern der logarithmierte Preis geschätzt wird.





Machine Learning

Schema eines (supervised) ML-Modells:



Machine Learning

Grundsätzlich werden die Daten in drei Teile aufgeteilt:

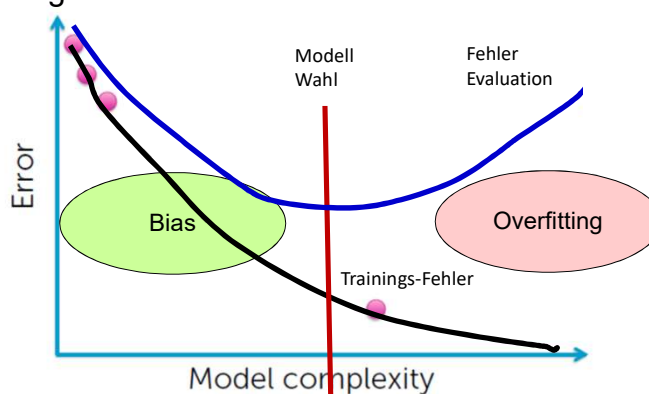
	Training set	Validation set	Test set
	80%	10%	10%
	50%	25%	25%

- Training Set: Auf diesen Daten werden verschiedene Modelle trainiert.
- Validation Set: Dient zum Festlegen der besten Modell-Parameter.
- Test Set: Mit diesen Daten wird der zukünftig zu erwartende Fehler geschätzt.



Machine Learning

Overfitting vs. Bias :



Machine Learning

Clustering: (Besprochenes Immobilien-Projekt)

Für eine qualitativ gute Schätzungen der Immobilienpreise müssen ähnliche Objekte zuerst (zu Cluster) zusammengefasst werden.

2 Beispiele:

Nr	Eigenschaften	Anteil %
1	Einfamilienhäuser	46.55
2	Mehrfamilienhäuser	12.55
3	Private Einstellhalle bis und mit 20 Plätzen	12.15
4	Restliche Datensätze	28.74

Nr	Eigenschaften	Anteil %
1	Einfamilienhäuser	35.51
2	Mehrfamilienhäuser	27.63
3	Autobox, Gartenhäuser, Velounterstand etc.	23.37
4	Büros, Lager, Mehrfamilienhäuser, Einstellhallen	13.10
5	Alte (1350-1600) Gebäude in Altstadt mit rund 4 Stockwerken	0.39



Deep Learning

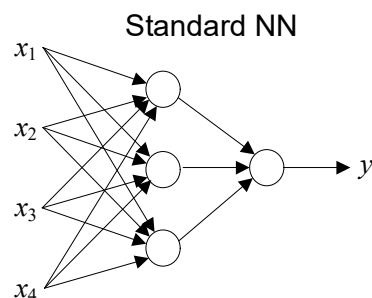
Die Basis aller Deep Learning Methoden sind Neuronale Netzwerke.

Wir unterscheiden verschiedene Typen von Neuronalen Netzwerken.

Beispiel:

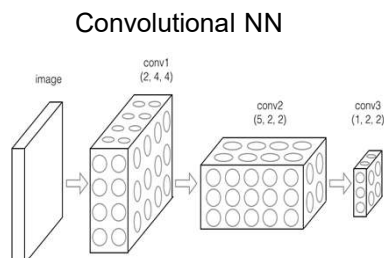
Input → Gebäudedaten

Output → Immobilien-Kennwerte



Deep Learning

Bei CNN lernen verschiedene «Filter», sukzessive komplexere Strukturen in Bildern zu erkennen.



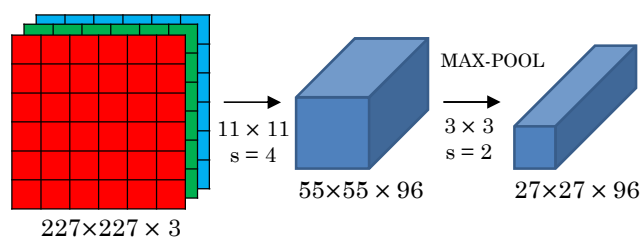
Beispiel:

Input → Bilder

Output → Klassifikation von Objekten



Deep Learning



Beispiel:

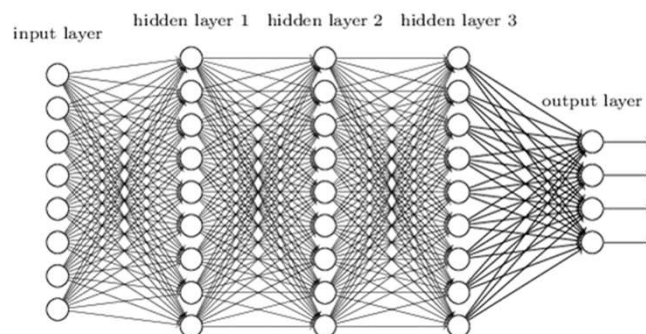
Input → Bilder

Output → Klassifikation von Objekten



Deep Learning

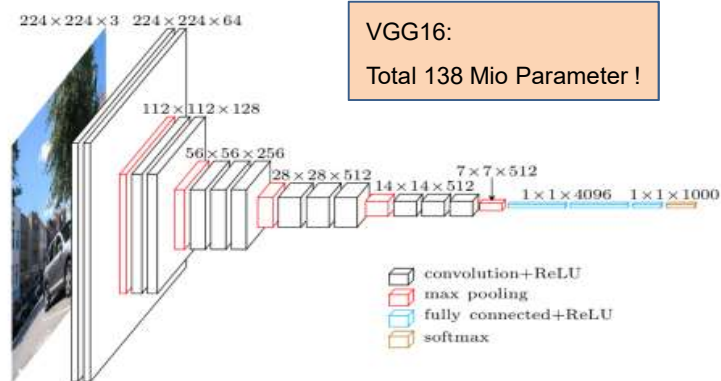
«Deep» bedeutet, sehr viele Layers :





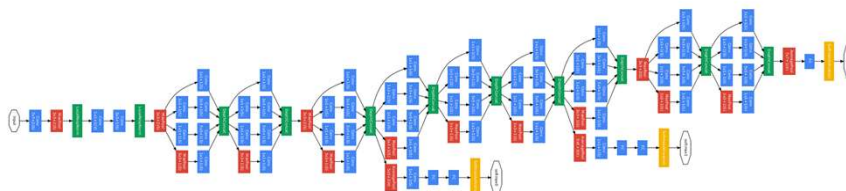
Deep Learning

Z.Bsp. 16 Hidden Layers:



Deep Learning

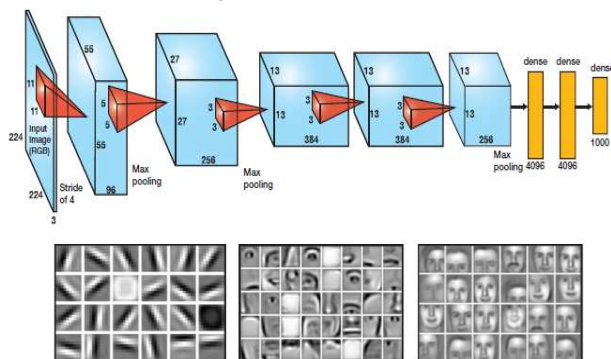
Immer «tiefer» Neuronale Netzwerke mit immer noch mehr Parameter sind schwierig zu optimieren und benötigen sehr grosse Rechenleistungen, wofür GPU's (Graphics Processing Unit) eingesetzt werden.





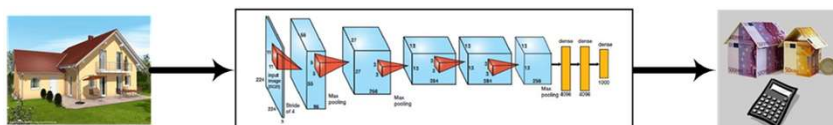
Deep Learning

Transfer Learning: Trainierte Netzwerke können teilweise übernommen und auf andere Anwendungen nach-trainiert werden.



Laufende Projekte

Schätzung von Immobilien-Kennwerten (Baukosten, Versicherungswerte, etc.) mit Hilfe von Deep Learning Methode und öffentlich zugänglichen Daten (GIS-Daten, Wohnlage, Lärmbelastung, etc.):



End-to-end Modell: Deep Neural Network für direkte Schätzung von Kennwerten



Laufende Projekte

Zillow (www.zillow.com) ist eine führende Real Estate Unternehmung mit einer Datenbasis von über 110 Mio. Gebäuden in den USA.

Mit Zestimate® verfügt Zillow über ein Tool um den Marktwert von Gebäuden zu schätzen.

Aktuell läuft ein (zweijähriger) Wettbewerb (dotiert mit 1.2 Mio \$) zur Verbesserung von Zestimate – unter den verbliebenen 100 Teams sind die 3 FHNW-Teams.

Datenbasis für das Training:

2.9 Mio. Gebäude mit 58 Features und 90'000 Transaktionen.

Daten für Testing:

Tatsächliche Verkaufspreise im September/Okttober 2018.



Fragen





University of Applied Sciences and Arts Northwestern Switzerland
School of Engineering



Besten Dank
für Ihre Aufmerksamkeit