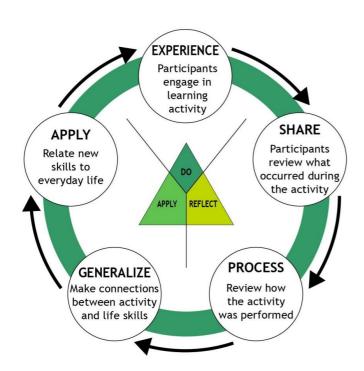


Inhalt

- Was ist (maschinelles) Lernen und wozu braucht man es?
- Was bisher geschah
- Was Sie (ggfs.) erwartet
- Sklearn Cheat Sheet
- Vergleich überwachtes, unüberwachtes, tiefes Lernen
- Überwachtes vs. unüberwachtes Lernen
- Semi-überwachtes Lernen
- Reinforcement Learning
- Ensemble Learning
- Ziele des Kurses
- Wichtige Anmerkung

Was ist Lernen?

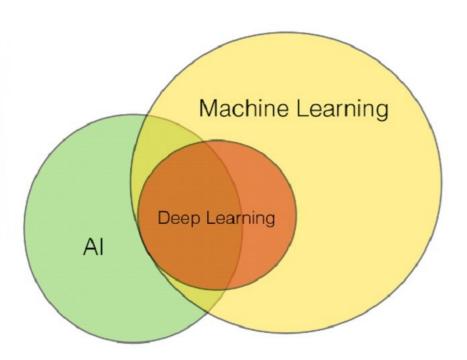
- Wikipedia: "Erwerb von Fähigkeiten"
- Verhaltensänderung durch Erfahrung
 - Dynamische Adaption (= veränderliche Anpassung)



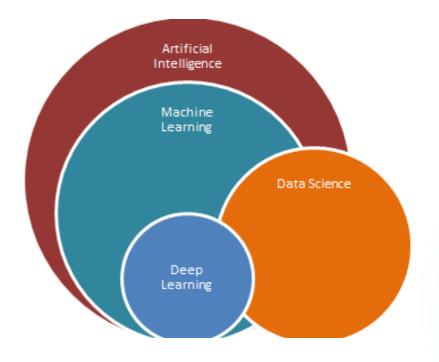
Experiential Learning Model

(Quelle: Uni Wisconsin-Madison)

Was ist Maschinelles Lernen?

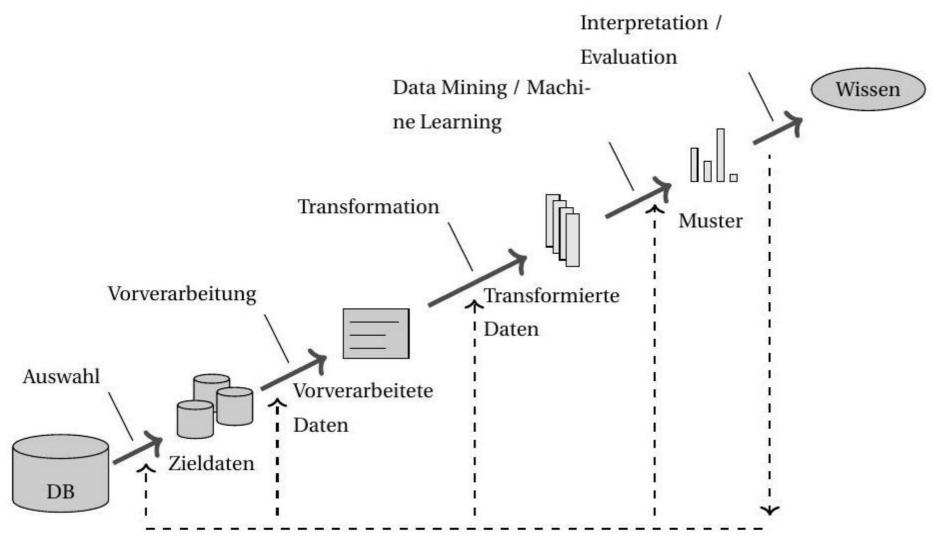


Quelle: Wikipedia



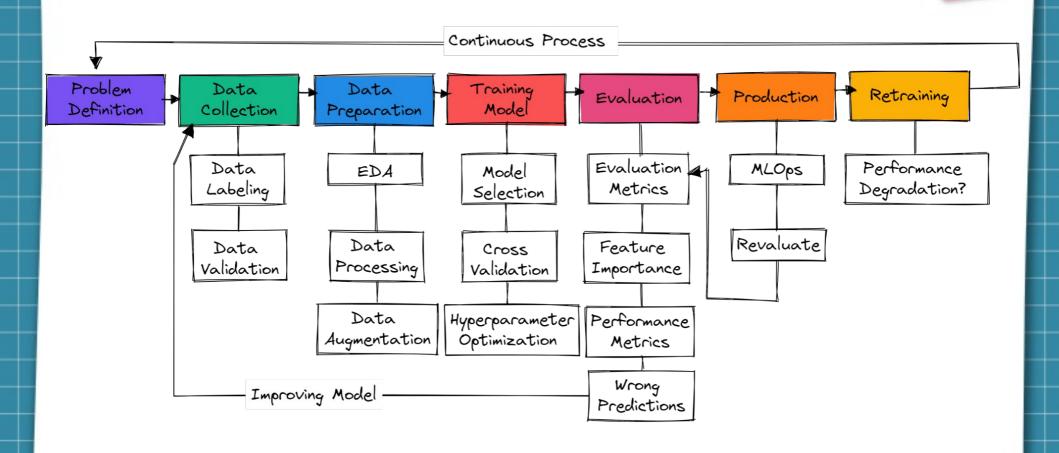
Quelle: medium.com

Knowledge Discovery in DBs



Quelle: Frochte, S. 17

How to start an ML-Project



Source: https://www.datacamp.com/blog/machine-learning-projects-for-all-levels

Was bisher geschah







BASIC REGRESSION









Target variable is categorical



human builds model based on input / output

human input, machine output human utilizes if satisfactory

human input, machine output human reward/punish, cycle continues

CLUSTER ANALYSIS





Similar datum into groups based on centroids







covariance. Elliptical Envelope() Finding outliers through grouping

CLASSIFICATION





Complex relationships. Prone to overfitting Basically magic.





neighbors.KNeighborsClassifier()

Group membership based on proximity





tree.DecisionTreeClassifier()

If/then/else. Non-contiguous data. Can also be regression.





ensemble.RandomForestClassifier() Find best split randomly Can also be regression





svm.SVC() svm.LinearSVC()

Maximum margin classifier. Fundamental Data Science algorithm





GaussianNB() MultinominalNB() Berno

Updating knowledge step by step with new info

FEATURE REDUCTION

T-DISTRIB STOCHASTIC NEIB EMBEDDING



manifold.TSNE()

Visual high dimensional data. Convert similarity to joint probabilities

PRINCIPLE COMPONENT ANALYSIS



decomposition.PCA()

Distill feature space into components that describe greatest variance

CANONICAL **CORRELATION ANALYSIS**



decomposition.CCA()

Making sense of cross-correlation matrices

LINEAR **DISCRIMINANT ANALYSIS**



Linear combination of features that separates classes

OTHER IMPORTANT CONCEPTS

BIAS VARIANCE TRADEOFF

UNDERFITTING / OVERFITTING

INERTIA

ACCURACY FUNCTION

PRECISION FUNCTION

SPECIFICITY FUNCTION

SENSITIVITY FUNCTION

Was Sie erwartet







BASIC REGRESSION





linear model.LinearRegression()

Lots of numerical data





Target variable is categorical

human builds model based on input / output

human input, machine output human utilizes if satisfactory

human input, machine output human reward/punish, cycle continues

CLUSTER ANALYSIS



Similar datum into groups based on centroids



covariance.EllipticalEnvelope()

Finding outliers through grouping

CLASSIFICATION





neural_network,MLPClassifier()

Complex relationships. Prone to overfitting Basically magic.







neighbors.KNeighborsClassifier()

Group membership based on proximity





tree.DecisionTreeClassifier()

If/then/else. Non-contiguous data. Can also be regression.





Find best split randomly Can also be regression







Maximum margin classifier. Fundamental Data Science algorithm





GaussianNB() MultinominalNB()

Updating knowledge step by step with new info

FEATURE REDUCTION

T-DISTRIB STOCHASTIC NEIB EMBEDDING



manifold.TSNE()

Visual high dimensional data. Convert similarity to joint probabilities

PRINCIPLE COMPONENT ANALYSIS





decomposition.PCA()

Distill feature space into components that describe greatest variance

CANONICAL **CORRELATION ANALYSIS**



decomposition.CCA()

Making sense of cross-correlation matrices

LINEAR **DISCRIMINANT ANALYSIS**





Linear combination of features that separates classes

OTHER IMPORTANT CONCEPTS



BIAS VARIANCE TRADEOFF

UNDERFITTING / OVERFITTING

INERTIA

ACCURACY FUNCTION

PRECISION FUNCTION

SPECIFICITY FUNCTION

SENSITIVITY FUNCTION

Originally Created by Emily Barry. See original here.

Was Sie ggfs. erwartet







BASIC REGRESSION





linear model.LinearRegression()

Lots of numerical data





Target variable is categorical

human builds model based on input / output

human input, machine output human utilizes if satisfactory

human input, machine output human reward/punish, cycle continues

CLUSTER ANALYSIS





cluster.KMeans()

Similar datum into groups based on centroids











covariance. Elliptical Envelope()

Finding outliers through grouping

CLASSIFICATION





neural network.MLPClassifier()

Complex relationships. Prone to overfitting Basically magic.





neighbors.KNeighborsClassifier()

Group membership based on proximity





tree.DecisionTreeClassifier()

If/then/else. Non-contiguous data. Can also be regression.





ensemble.RandomForestClassifier() Find best split randomly Can also be regression





svm.SVC() svm.LinearSVC()

Maximum margin classifier. Fundamental Data Science algorithm





GaussianNB() MultinominalNB() Berno

Updating knowledge step by step with new info

FEATURE REDUCTION

T-DISTRIB STOCHASTIC NEIB EMBEDDING



manifold.TSNE()

Visual high dimensional data. Convert similarity to joint probabilities

PRINCIPLE COMPONENT ANALYSIS



decomposition.PCA()

Distill feature space into components that describe greatest variance

CANONICAL **CORRELATION ANALYSIS**



decomposition.CCA()

Making sense of cross-correlation matrices

LINEAR **DISCRIMINANT ANALYSIS**



Linear combination of features that separates classes

OTHER IMPORTANT CONCEPTS

BIAS VARIANCE TRADEOFF

UNDERFITTING / OVERFITTING

INERTIA

ACCURACY FUNCTION

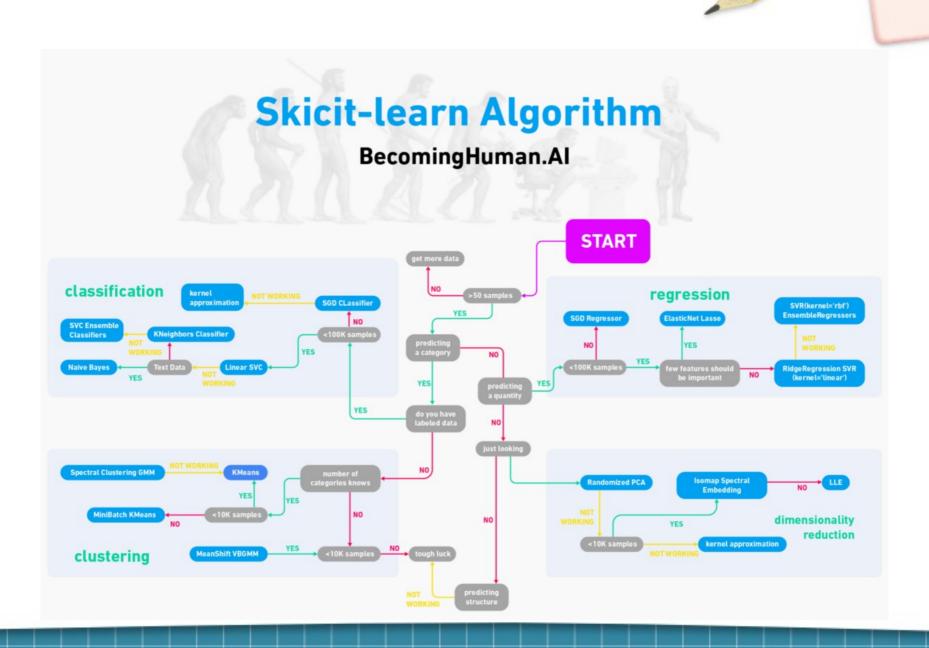
PRECISION FUNCTION

SPECIFICITY FUNCTION

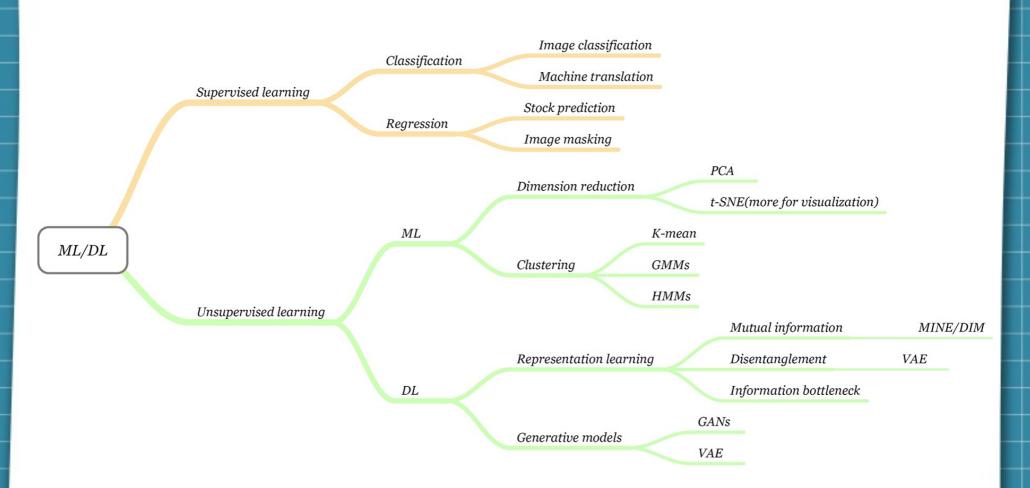
SENSITIVITY FUNCTION

Originally Created by Emily Barry. See original here.

Sklearn Cheat Sheet



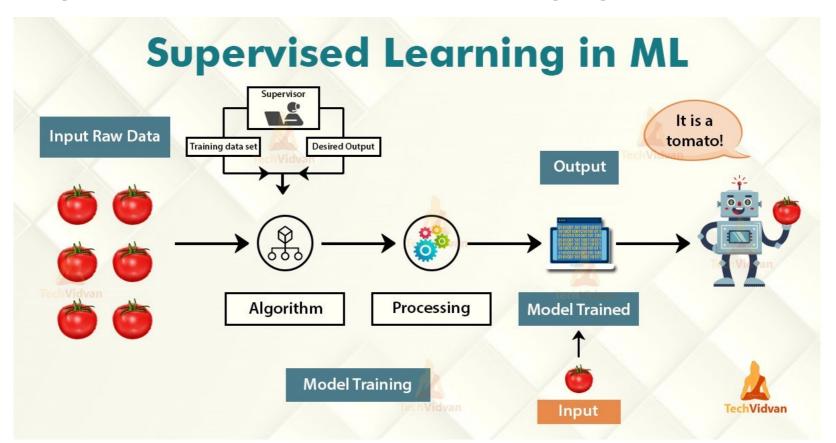
Vergleich SL/UL/DL



Quelle: medium.com

Überwachtes Lernen

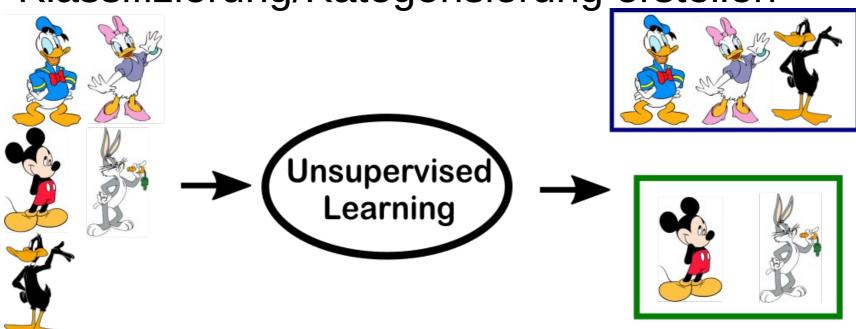
- Supervised Learning
- Input Data und Labels sind gegeben



Unüberwachtes Lernen

- Unsupervised Learning
- Keine Labels sind gegeben

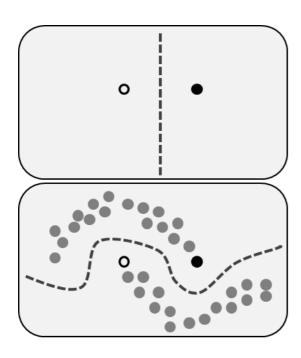
 Der Algorithmus muss selbst eine Klassifizierung/Kategorisierung erstellen



Quelle: medium.com

Semi-überwachtes Lernen

- Mischform aus überwachtem und unüberwachtem Lernen
- Teils fehlende, verrauschte oder ungenaue Zielvariablen



Quelle: Wikipedia

Reinforcement Learning

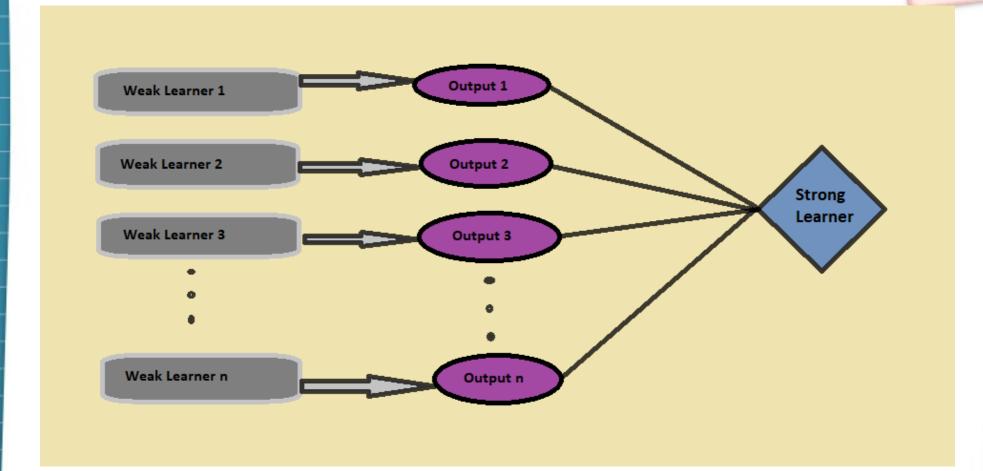
- Dynamische Interaktion des Algorithmus mit seiner Umgebung
- Pos. / neg. Feedback → Optimierung



Autonomes Fahren

Open AI Five

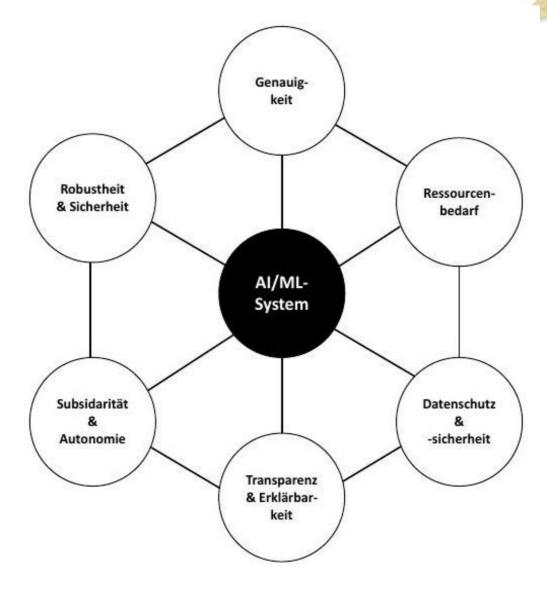
Ensemble Learning



Ziele des Kurses

- Durchführung von geeignetem Preprocessing
 - Missing Value Handling
 - Outlier Detection
 - Standardization / Normalization
 - Data Balance
- Entwurf und Implementierung eines Ensembles aus mind. zwei geeigneten schwachen Lernern
 - Voting / Bagging / Boosting / Stacking
- Ggfs. automatisierte Hyperparametersuche

Wichtige Anmerkung



Anforderungshexagon ML, Quelle: Frochte, S. 29

Danke für die Aufmerksamkeit.

