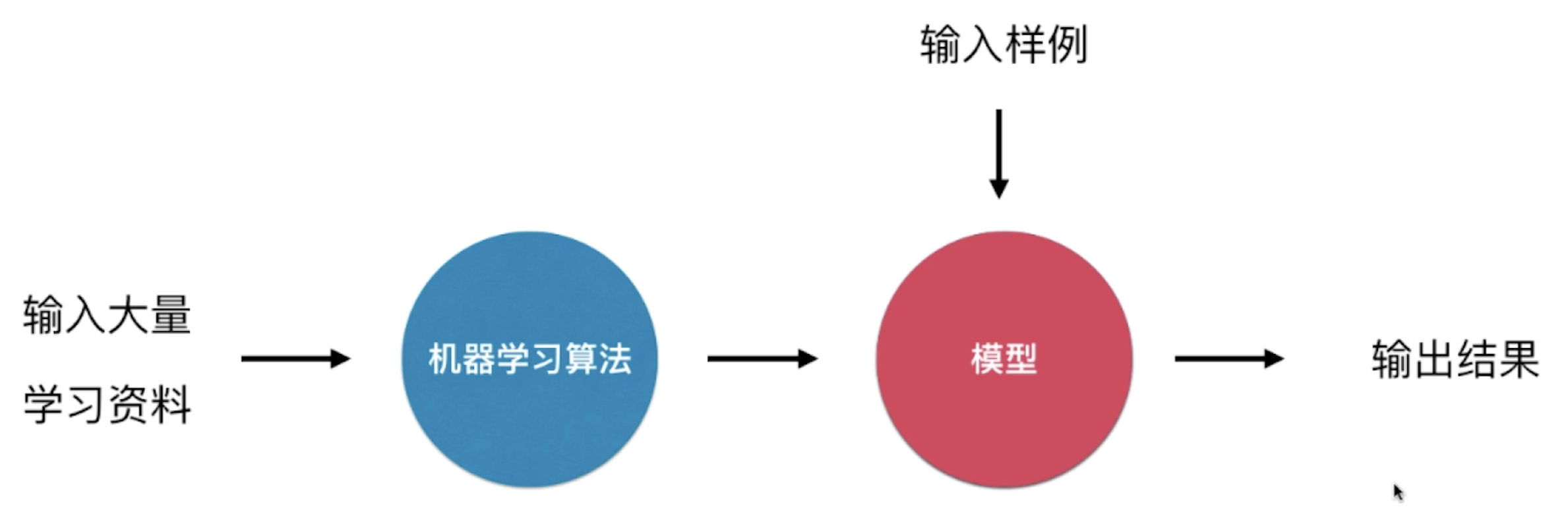
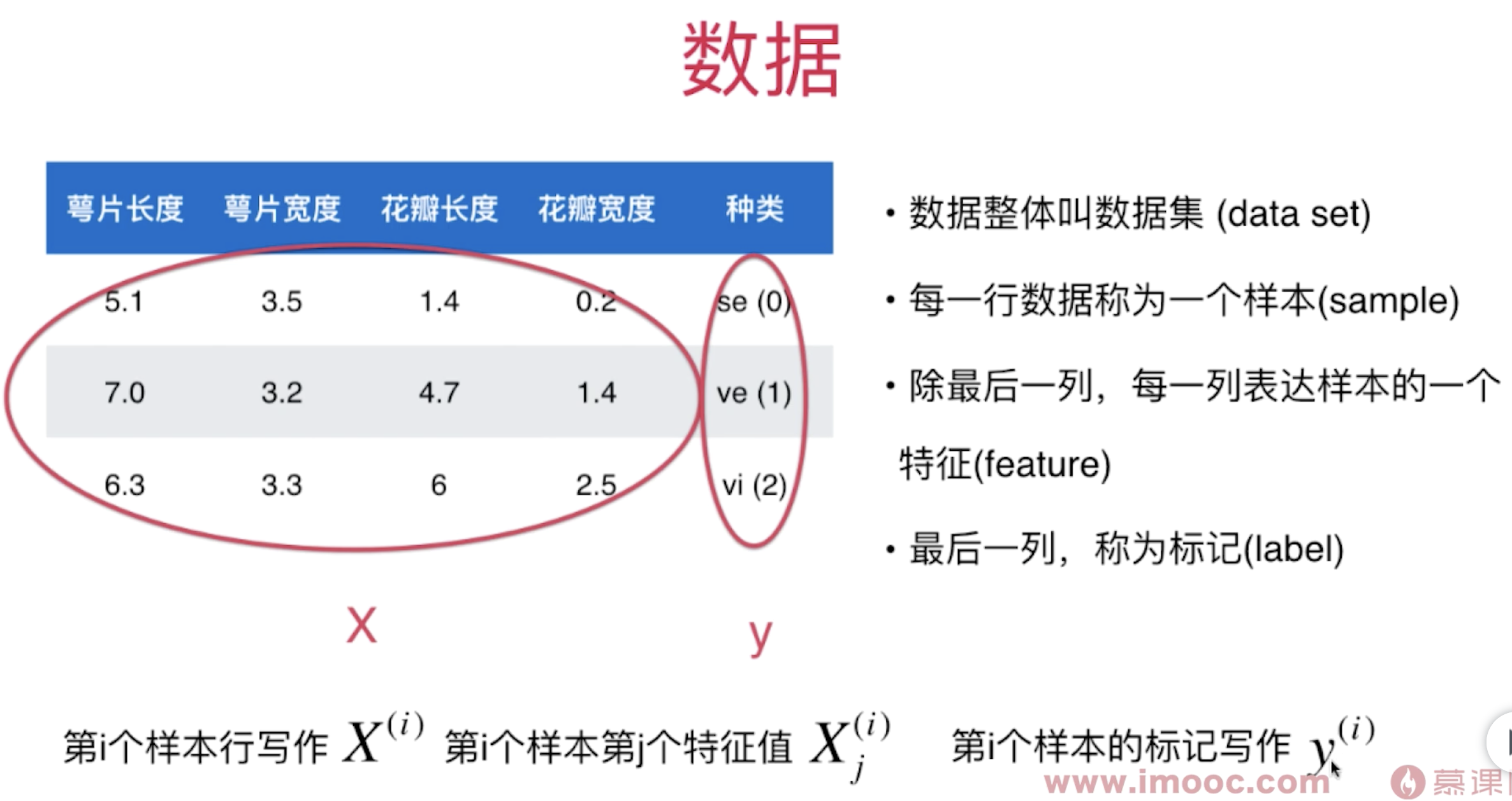
**Einführung Maschine Learning**

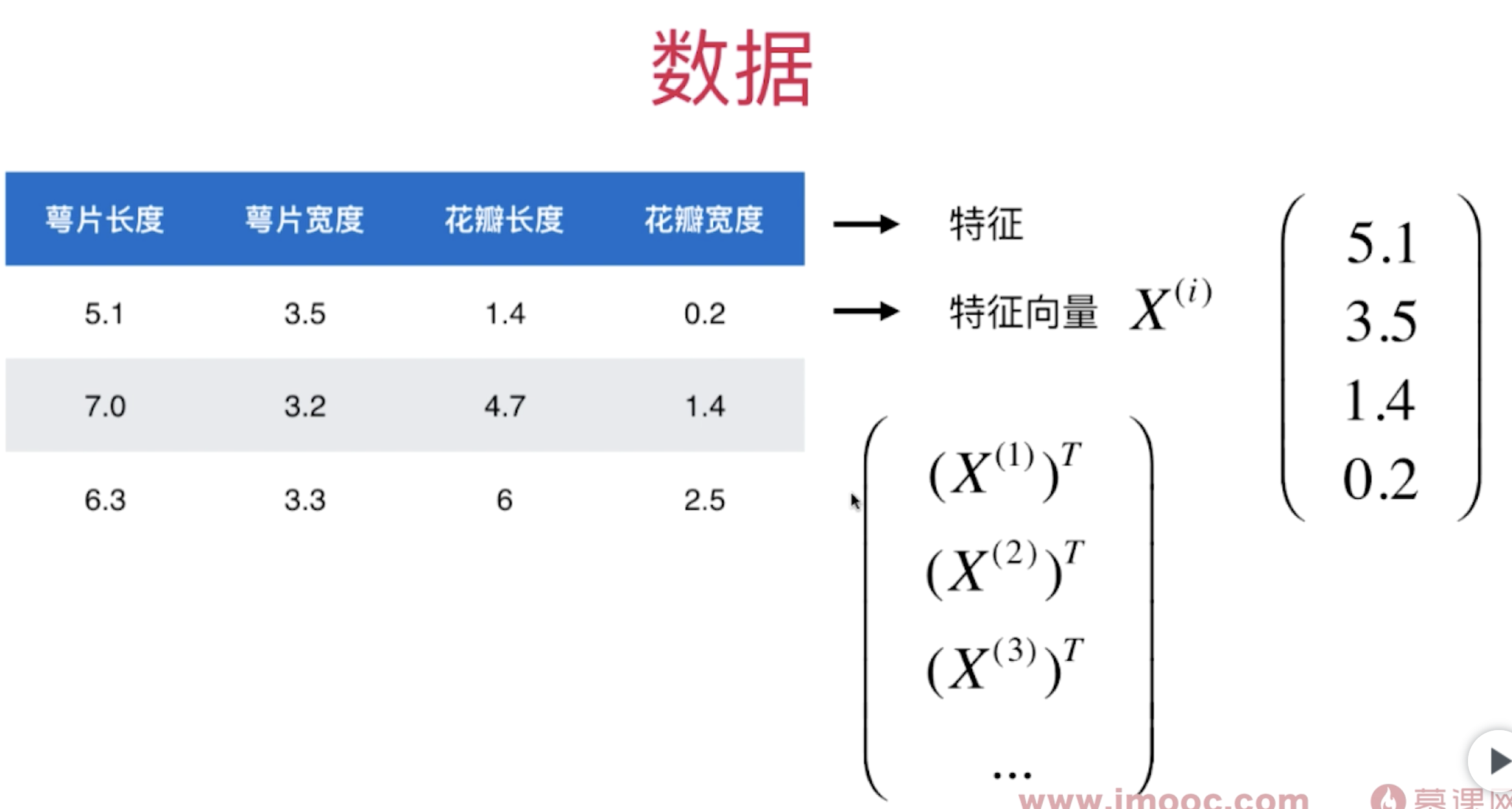
1. Was ist ML (Machine Learning)?

Geben viele Lernmaterialien ein, um ein Modell zu trainieren.



Fit (拟合)模型最终可理解为一个具体f(x), predict (预测)输出为集合则为分类问题, 输出为值则为回归问题





T (Matrix transpose)

### 2. Feature Space (Extrahiere Features aus data set, um eine n-Dimension zu bilden) (Segmentierung im Merkmalsraum (Feature space))

### 

**Abbildung 1 Iris Data Set (2 Features)**

### 3. Classification, Regression (two significant tasks) (supervised learning)

### Label classification(标签分类), multiple classification(多分类)

### 4. Supervised learning, Unsupervised learning, Semi-supervised learning, Reinforcement learning.

### Supervised learning: Die Trainingsdaten der Maschine haben “Labels” oder “Antworten”

### B.s. 图像拥有了标定信息

### Unsupervised learning: Die Trainingsdaten der Maschine haben keine „Labels“ oder „Antworten“.

### B.s. k-means, points have no external classification. 聚类分析(电商对用户进行划分)

### 数据进行降维处理(特征提取: 信用卡等级和人体身高无关, 特征压缩:PCA)

### PCA: 特征相关性很强,压缩成一条直线 尽量少损失数据的情况下,将高维度的特征向量压缩成低维的特征向量 提高运行效率 方便可视化(高维空间难以理解) 异常检测

### 

### Semi-supervised learning: Ein Teil der Trainingsdaten enthält Labels oder Antworten, der andere Teil nicht. 通常先用非监督学习对数据做处理, 之后用监督学习做模型训练和预测.

### Reinforcement learning: Erlernen optimaler Strategien, KI lernt in einer Simulationsumgebung durch Try and Error und ein Belohnungssystem eine Strategie zur Lösung eines Problems.

### 

### 

### 

### 5. Training -und Testdatensatz

### a). Das Ziel der Trainingsphase ist es, Modelle zu trainieren, die möglichst gut generalisieren, d.h. Modelle, die richtige Vorhersagen auf neuen Samples zu treffen.

### b). Übliche Aufteilung in einen Trainings- und einen Testdatensatz: 70%:30% oder 80%:20%

### c). Möglichkeiten der Aufteilung

### Zufällig, manuell, Stratifizierte Stichprobe

### 

### 6. Hyperparameter

### a). Parameter des Algorithmus, nicht des Modells (Modellparameter).

### b). Diese werden nicht gelernt, sondern müssen vor dem Training gesetzt werden und bleiben während des Trainings konstant.

### c). Die meisten Algorithmen besitzen mindestens einen Hyperparameter.

### d). Hyperparameter sind für das resultierende Modell ausschlaggebend.

### e). Wahl von Hyperparametern ist oft nicht trivial, meist ist Tuning notwendig.

### 7. Daten

### a). Zeit investieren und explizit einplanen, um brauchbare Daten zu sammeln und diese gründlich aufzubereiten.

### b). Minderwertige Daten

### 1). Starkes Rauschen

### 2). Hoher Anteil des Outliers

### 3). Fehlende Werte

### 4). Unpassende Verteilung

### c). Viele Irrelevante Features

### 8. Transformation in einen geeigneten Feature Space

### One-hot-Encoding

### 