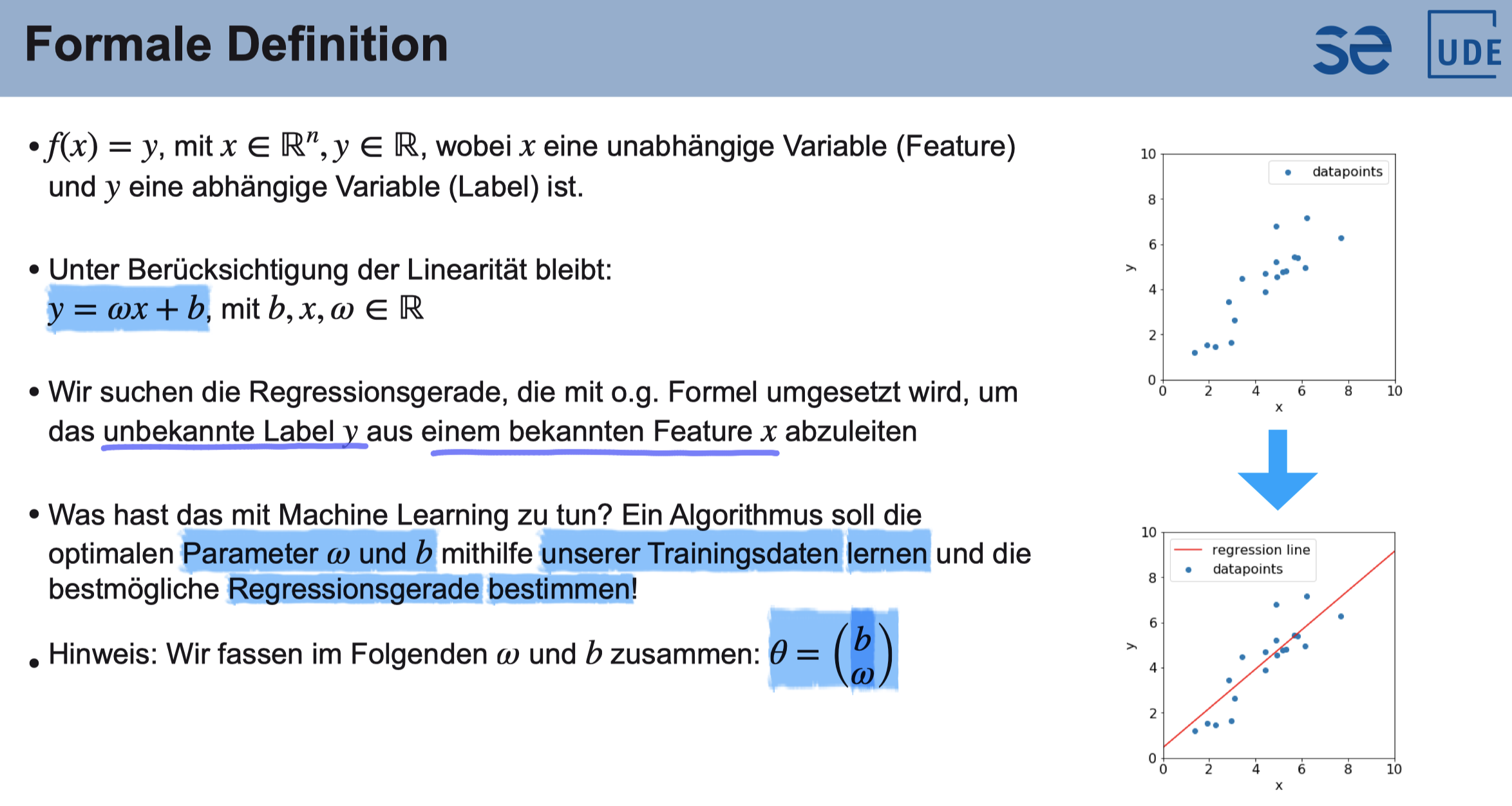
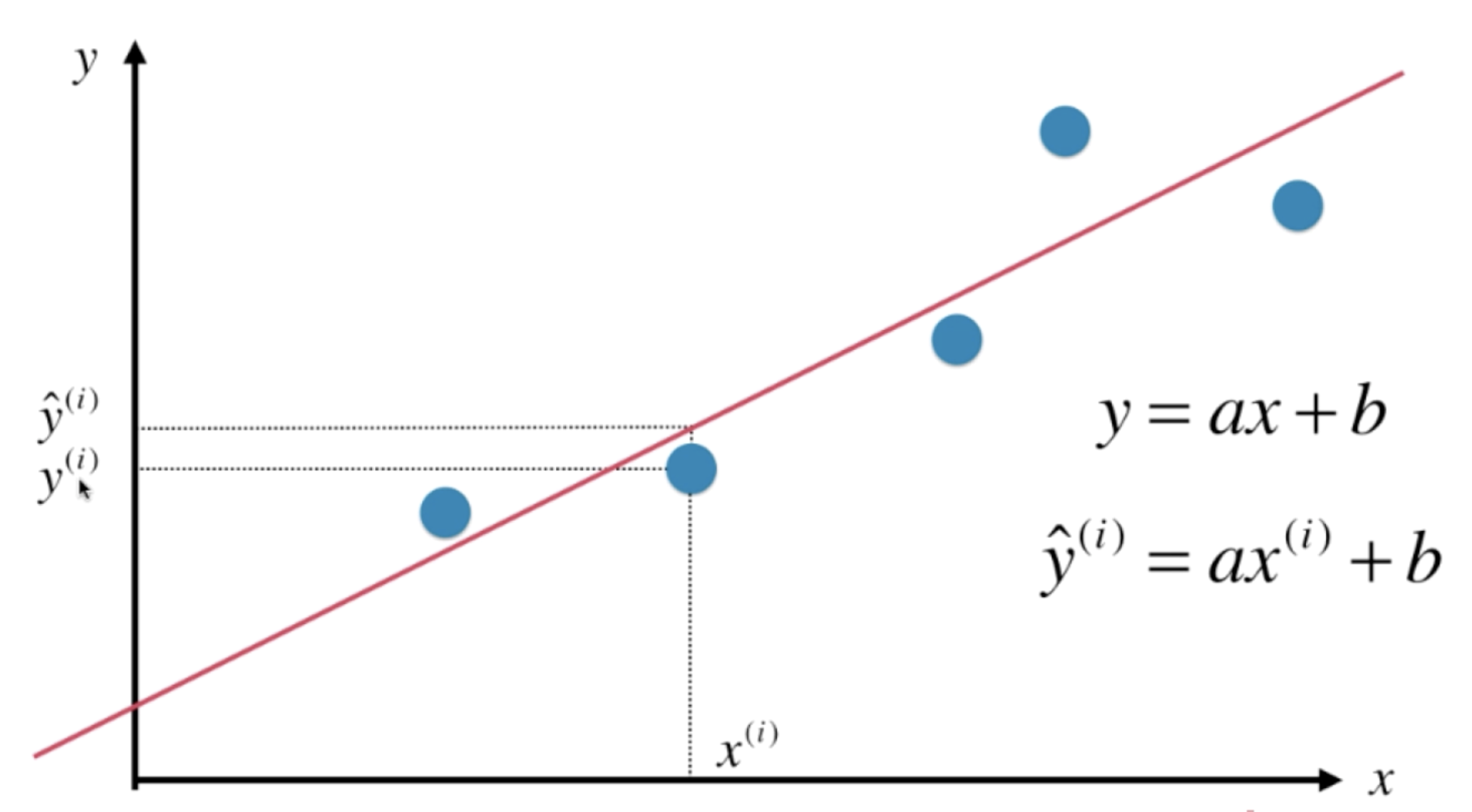
**2 linear Regression**

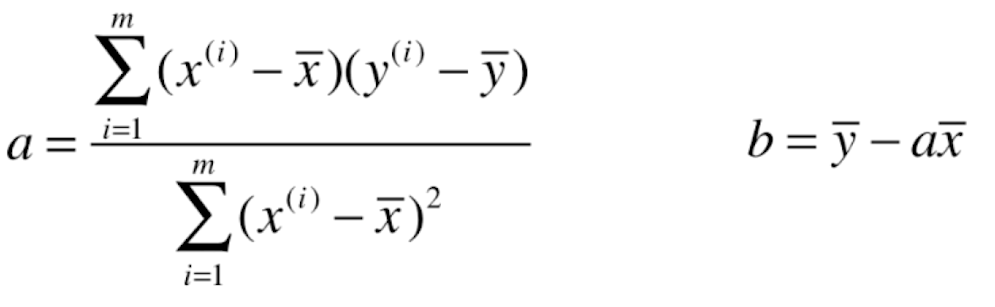
1. Formale Definition

X: Feature

Y: Label







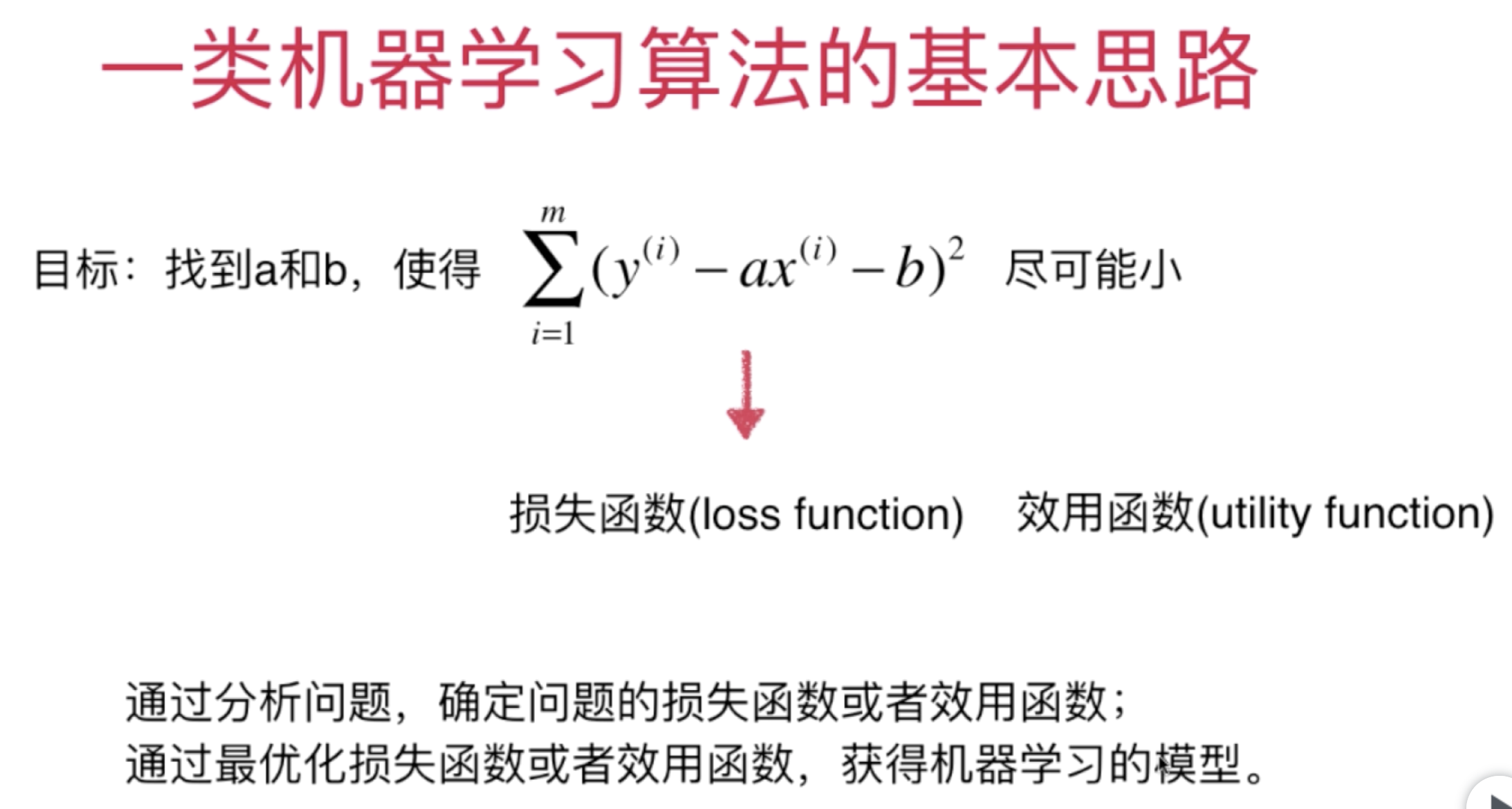
最小二乘法

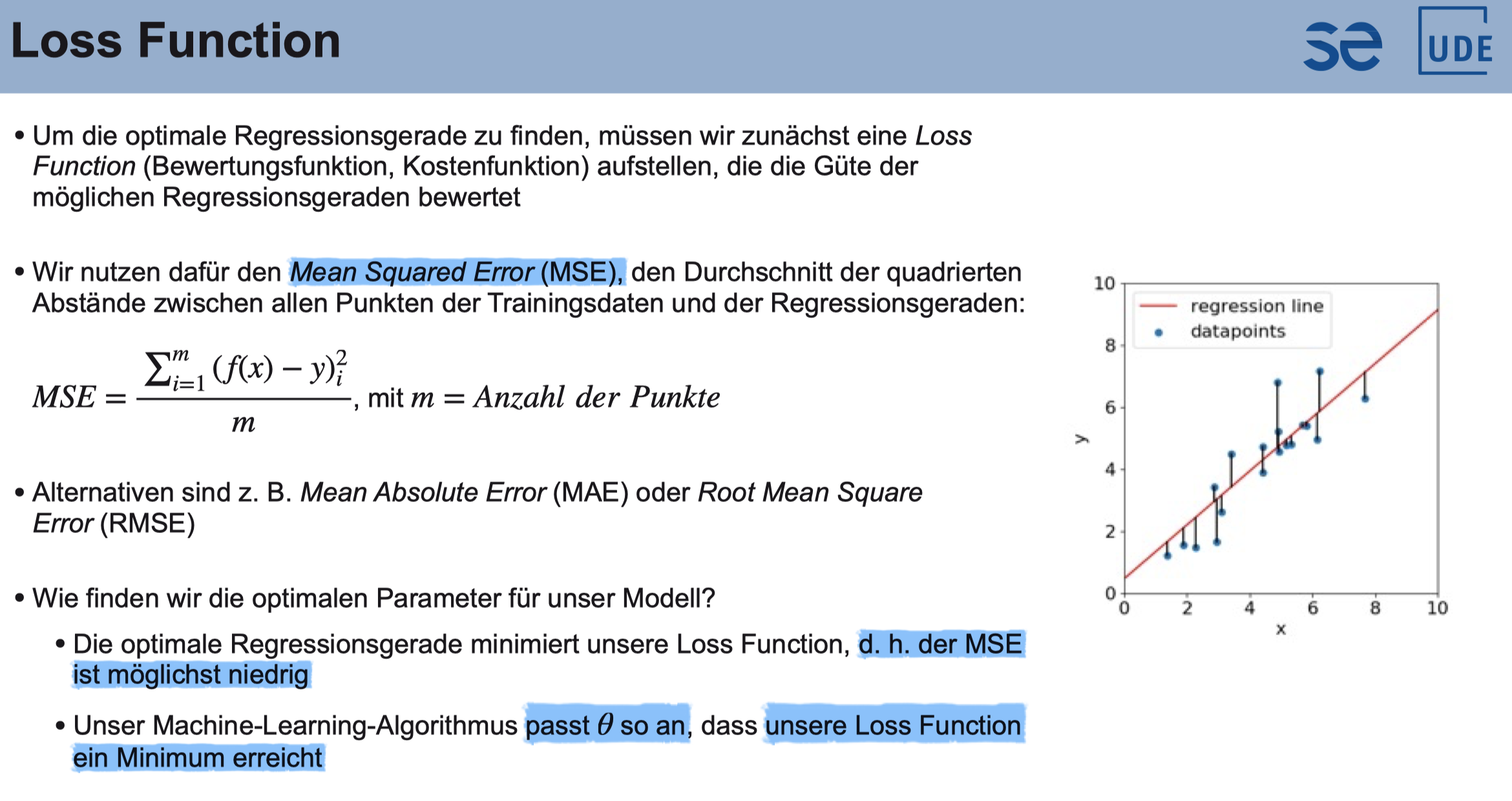
2. Metrics, Loss Funktion (preferably less) und Utility Funktion (preferably great)

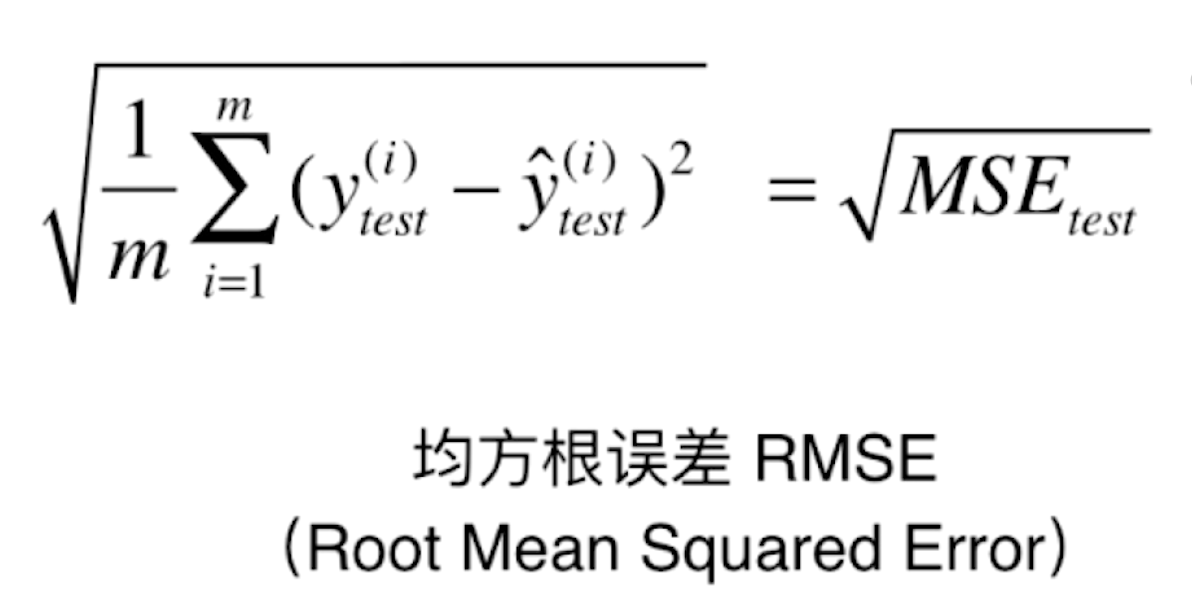
Loss funktion 尽可能小

Utility function 尽可能大

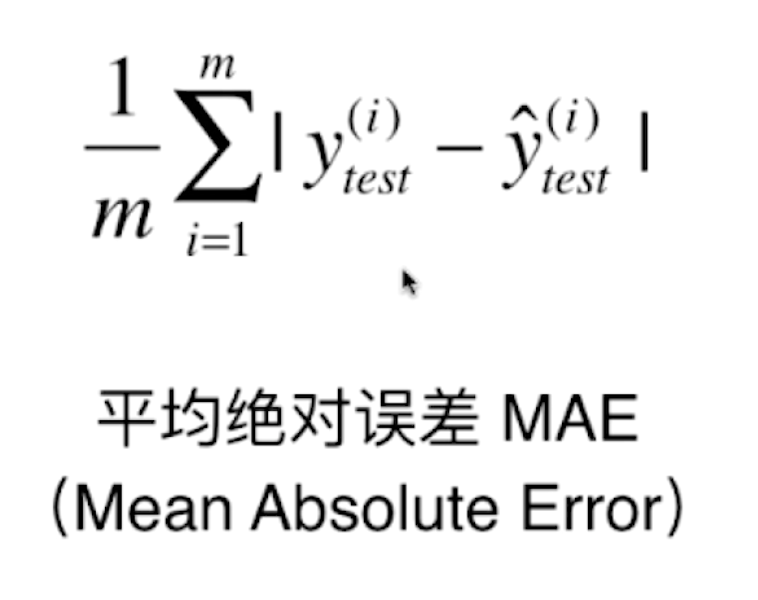
适用所有参数学习的方法, 先创建一个模型, 并且学习这个模型的参数



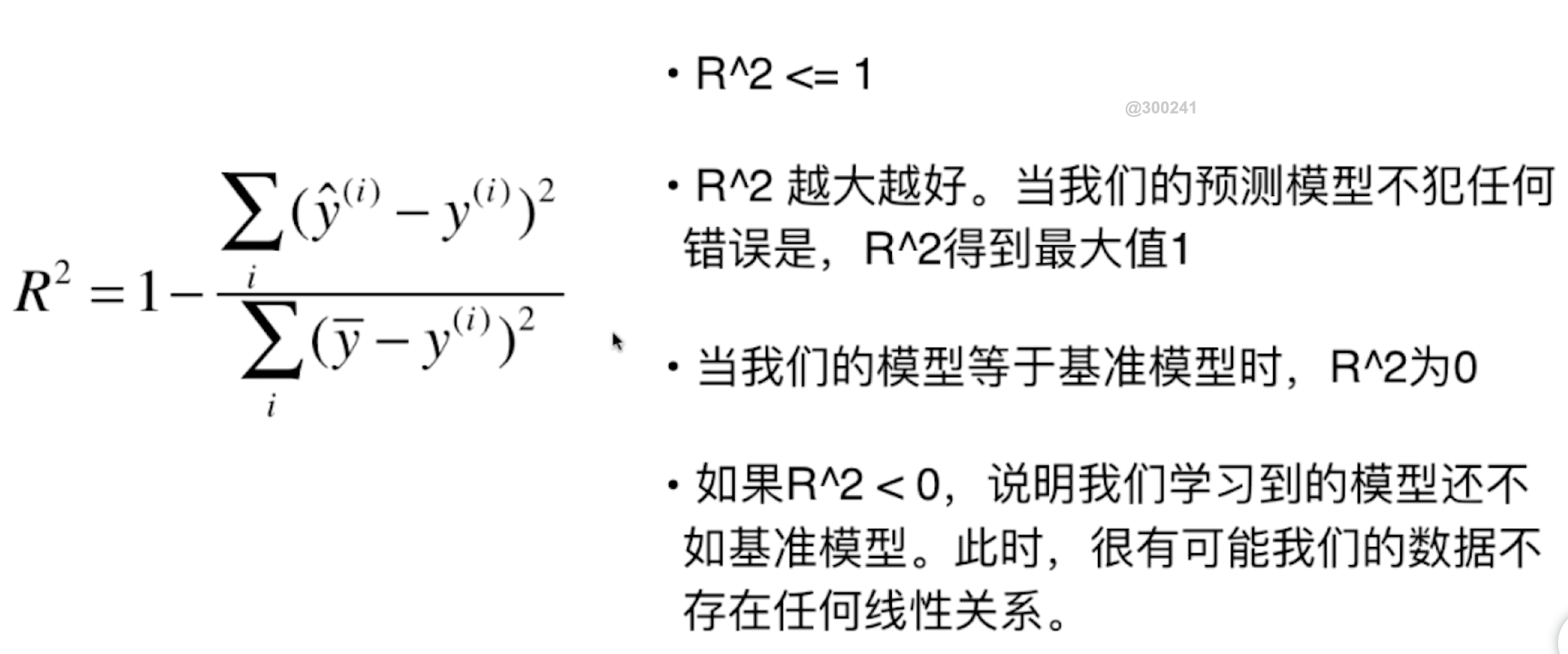




RMSE值更小意义更大

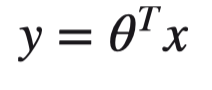


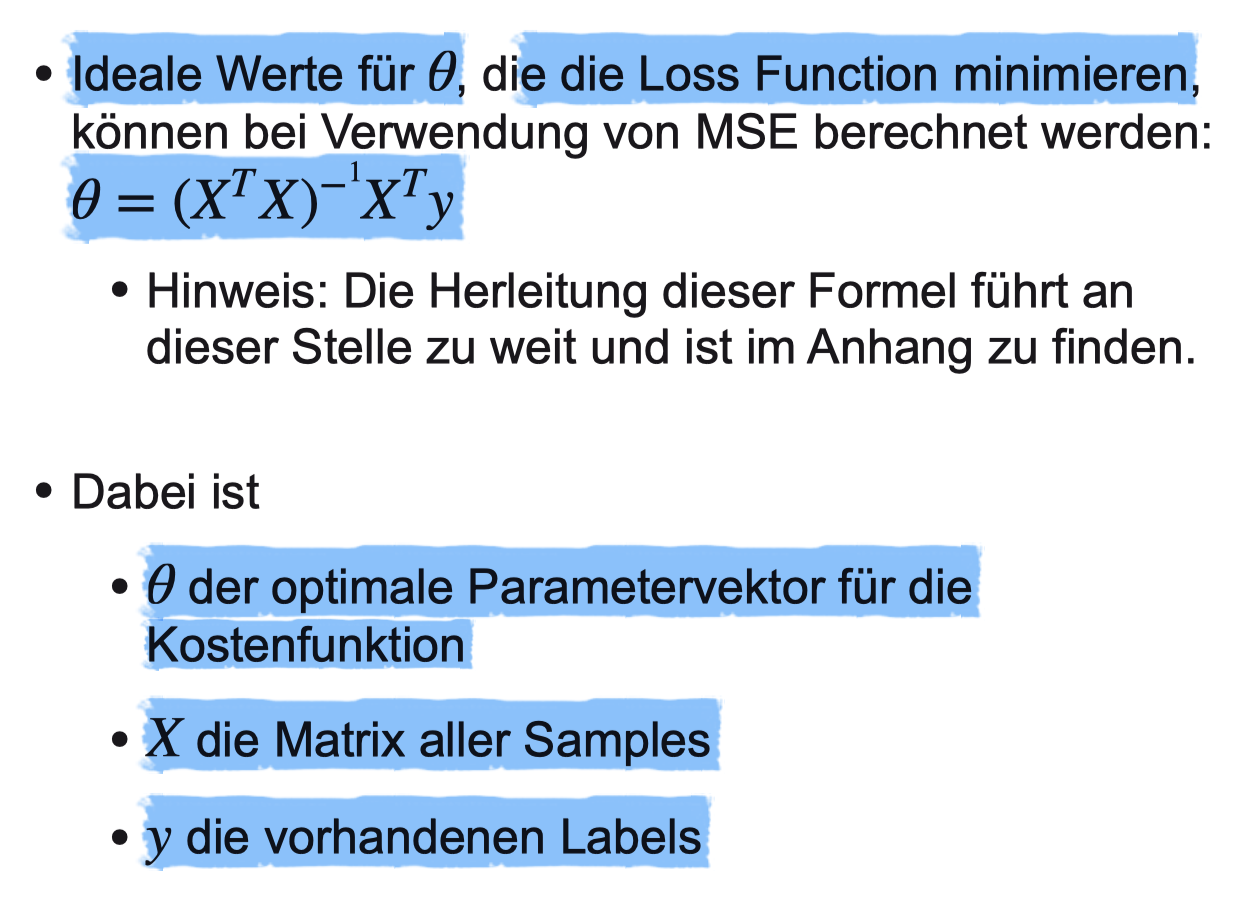
**The best metrics of LR: R squared**

****

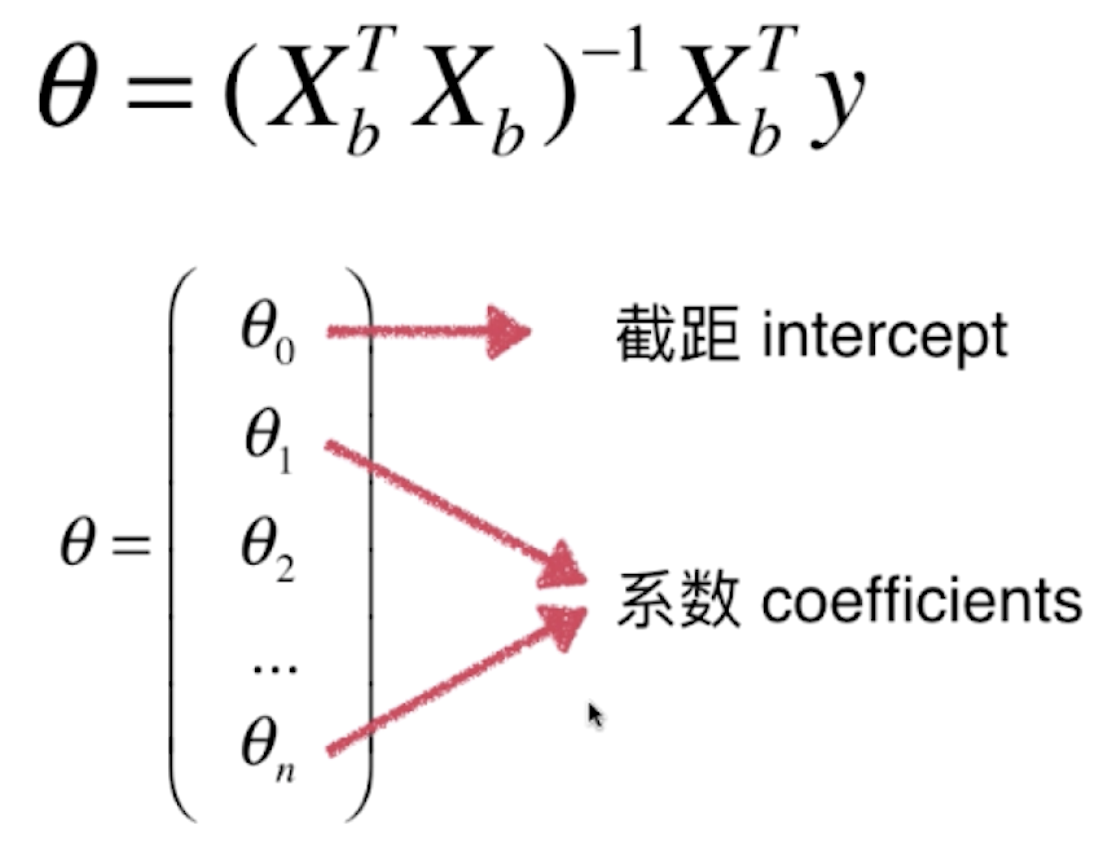
分母为baseline model基准模型 R squared = 1-MSE/Var

3. Multiple features

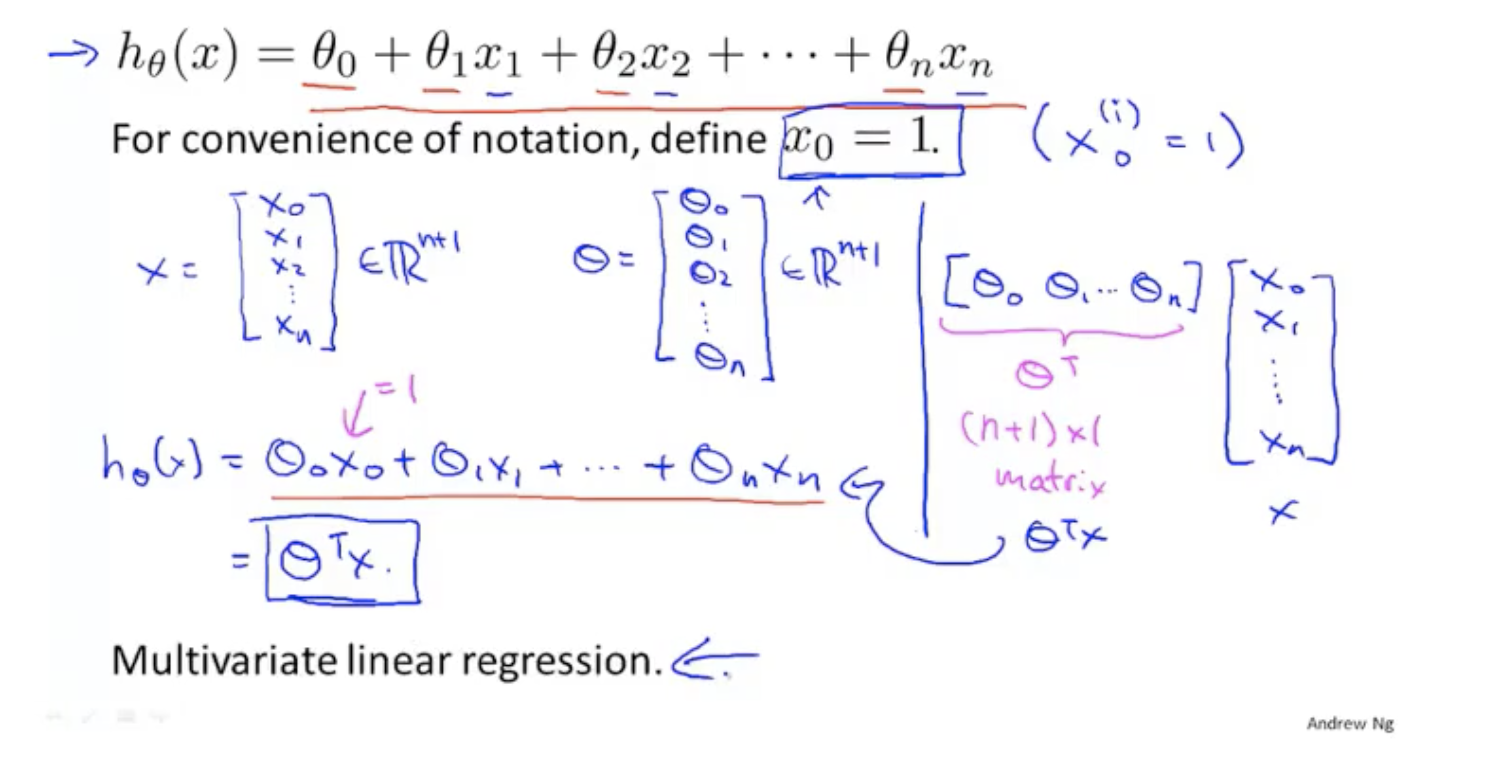
Normal Equation: 



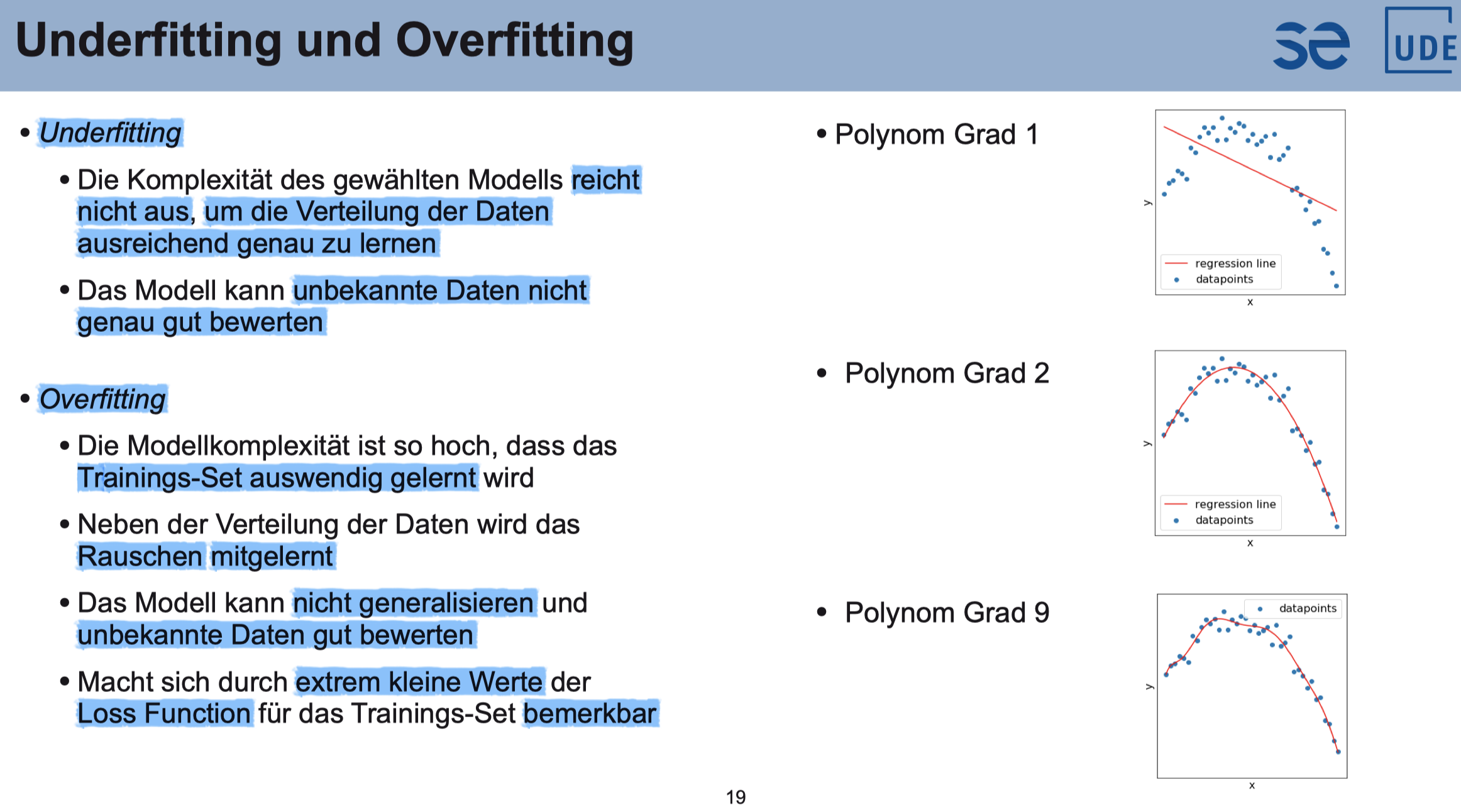
Interception, coefficient



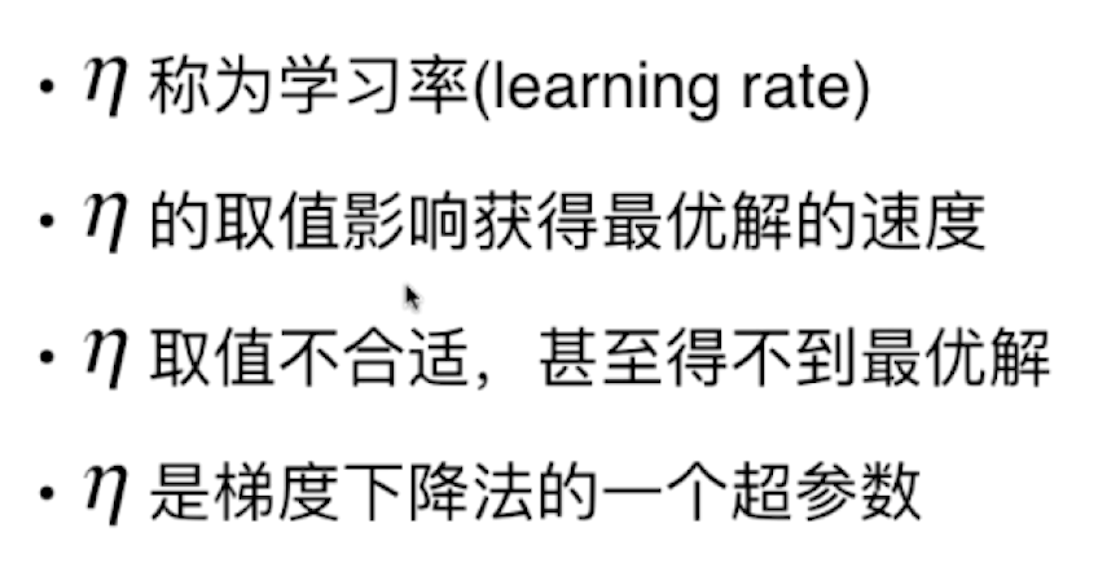
Theta : Loss function optimal mathematical solution

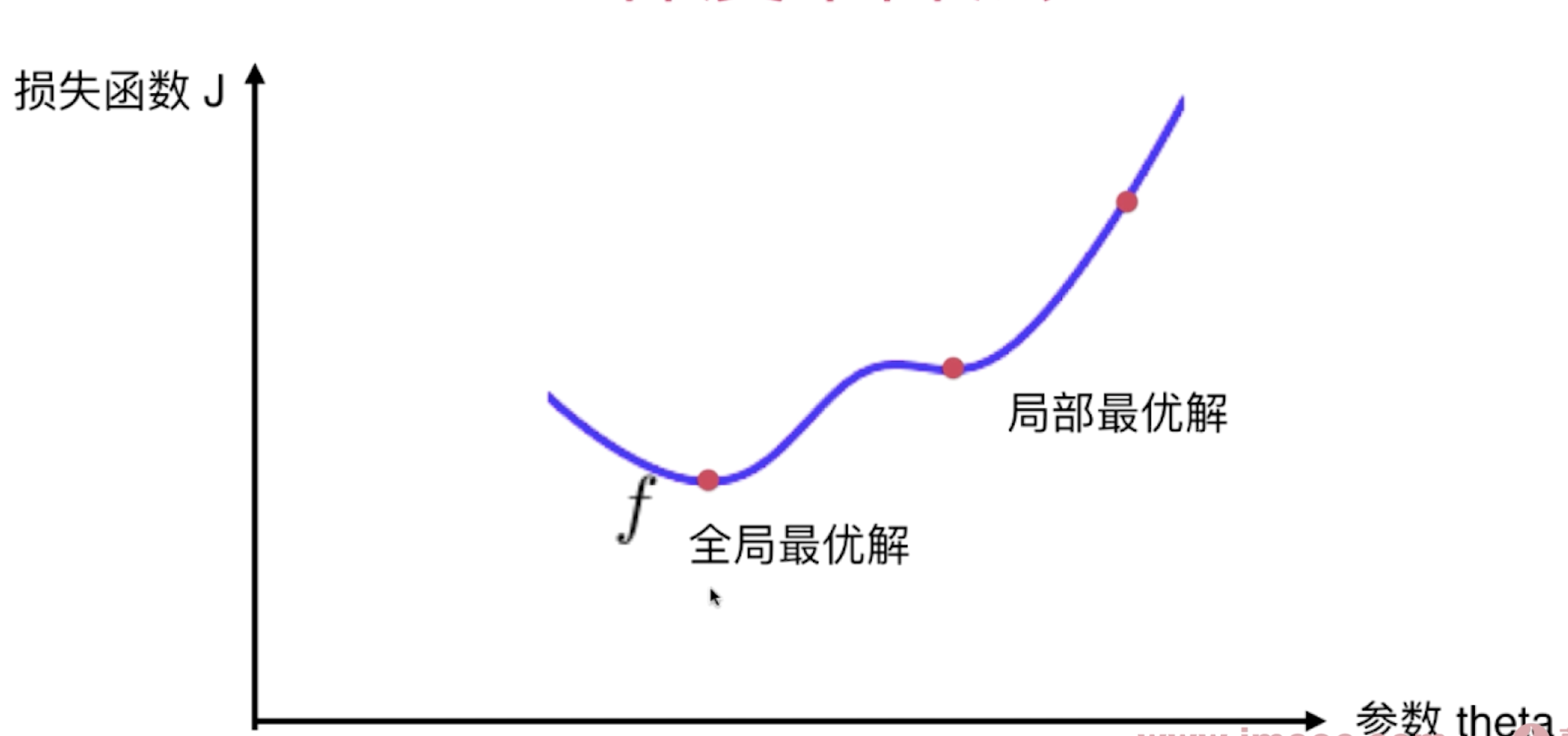


4. Underfitting und Overfitting



5. Gradient Descent: eine suchbasierte (**Der Extrempunkt ist nicht eindeutig, lokale optimale Lösung, globale optimale Lösung**) Optimierungsmethode, um eine Loss funktion zu minimieren.





Idee: Use GD many times and randomize the initial point

Gradient descent in Linear Regression:

