Herzlich Willkommen beim nächsten Video.

Heute behandeln wir den wohl kompliziertesten Teil der theoretischen Einführung in die Neuronalen Netzwerke. Denn wir haben noch gar nicht besprochen wie ein Allgemeines NN trainiert wird. Denn das Verfahren beim Perzeptron ist nicht allgemeingültig anwendbar.

Unser Ziel ist:

* Optimale Gewichte für w\_i bestimmen durch ableiten der Fehlerfunktion
* Rückschlüsse auf die einzelnen Gewichte ziehen und anhand des Gradienten diese dann mit einer gegebenen Lernrate verändern

Vorgehen:

* Gradientenabstieg: initialisiere die Gewichte zufällig, delta wij ist dann die Bewegung in die Richtung des neg. Gradienten
* Gradientenabstieg hängt stark von der Anfangsbedingung ab => verschiedene Startwerte benötigt

Eigenschaften:

* Es gibt keine optimale Lernrate
* Fehlerberechnung startet bei Output Neuronen
* Komplexität eines Updateschrittes ist linear
* Minima sind nicht eindeutig

Durchlauf:

* Mehrere Epochen und Epochen in Batches unterteilen
* Eine Epoche ist der Durchlauf des Gesamten Datensatzes
* Datensatz in Batches aufteilen, die einzelnen dann Trainieren und davon den Durchschnitt bilden und darauf dann Backpropagation ausführen und Gewichte anpassen
* Deshalb keiner normaler GD sondern Stochastic gradient descent (SGD)

Gibt noch sehr viele verschiedene Optimierugsverfahren, die auf SGD aufbauen!

Gibt verschiedene Error Funktionen die angewendet werden können!  
Initialisierung der zufälligen Gewichte anhand verschiedener Verteilungsfunktionen!

* Welche Verteilung gewählt werden soll, ist abhängig vom Datensatz (Erfordert Wissen über die Eigenschaften der Daten im Statistischen Sinne)