Tensor Wrold
人工智能讲座

深度学习课程

AI工程师讲座

问题分析 数据准备 特征工程 模型设计 总结回顾

目录

Part / 01

历史沿革

HISTARY

历史沿革

HISTARY

MP神经元

BP训练

Yann Lecun

1943

40年代

1974

90年代

1998

21世纪

Hebb学习
数学证明缺陷

SVM

深度神经网络

Part / 02

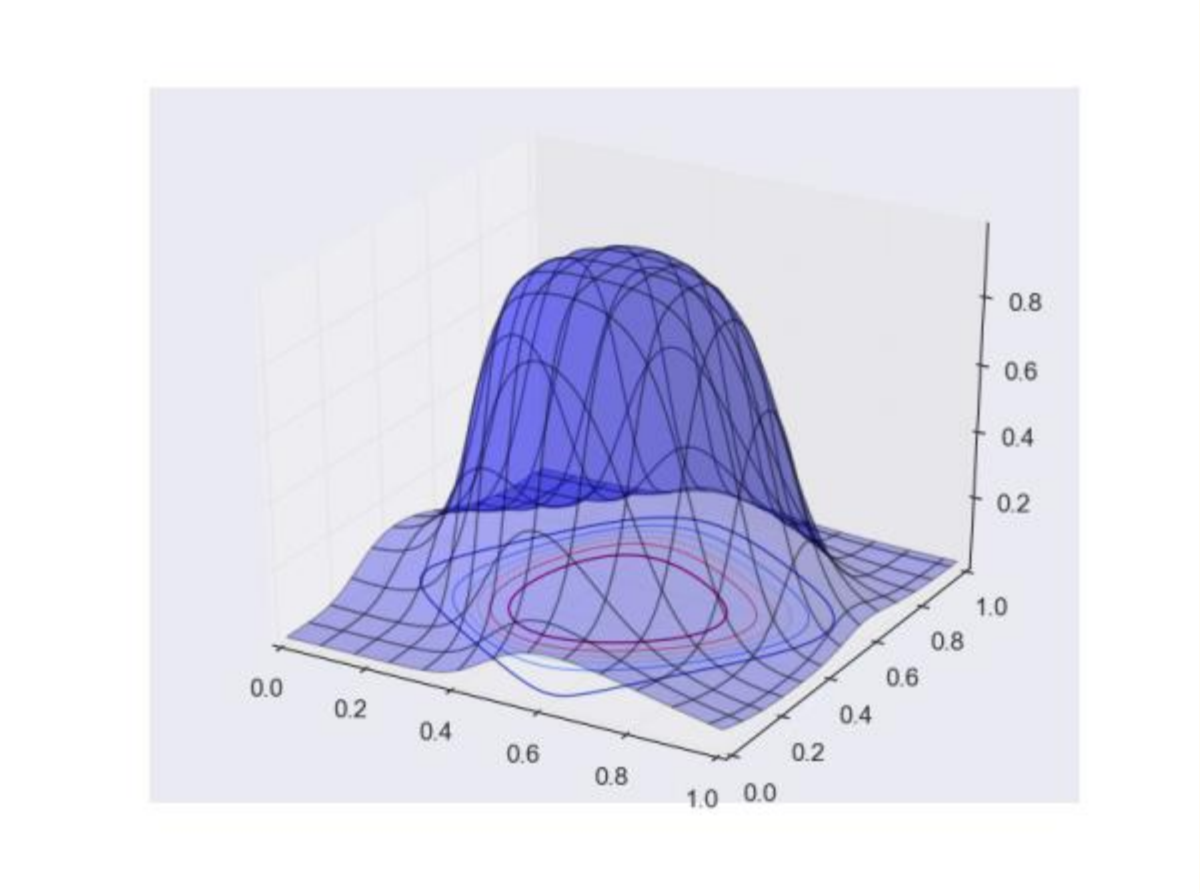
数学模型

MATHEMATICAL MODEL

对于前馈神经网络

$$\vec{y} = f(\vec{x})$$

直观理解
“多维曲面拟合”



对于前馈神经网络

$$\vec{y} = f(\vec{x})$$

单层

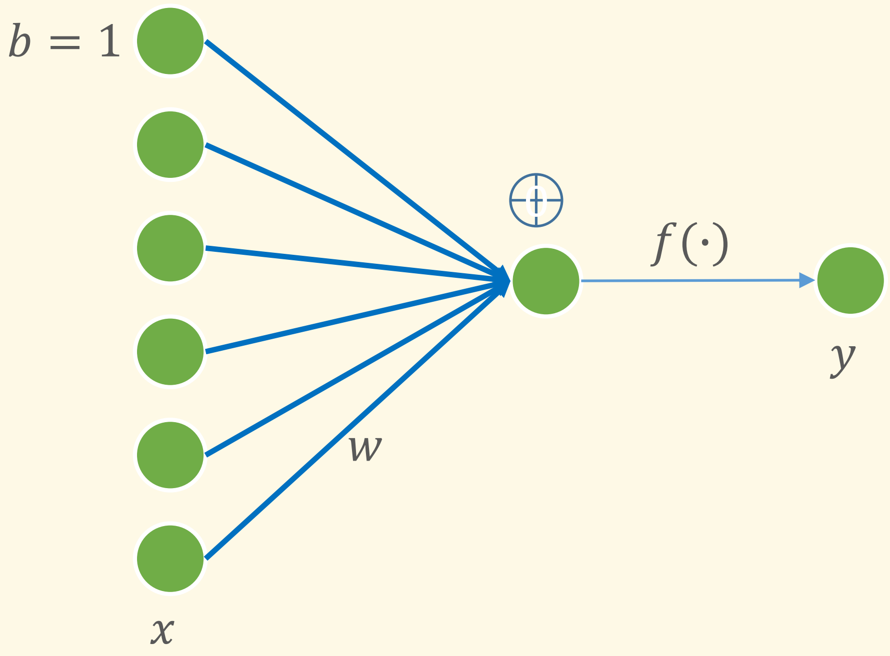
$$\vec{y} = f(W \cdot \vec{x})$$

$$f(\vec{x}) = Uint(\cdot)$$

最初激活函数

$$f(\vec{x}) = Sigmoid(\vec{x})$$
$$f(\vec{x}) = tanh(\vec{x})$$
$$f(\vec{x}) = ReLU(\vec{x})$$

常用激活函数



激活函数
存在梯度(一阶梯度)
对于训练很重要

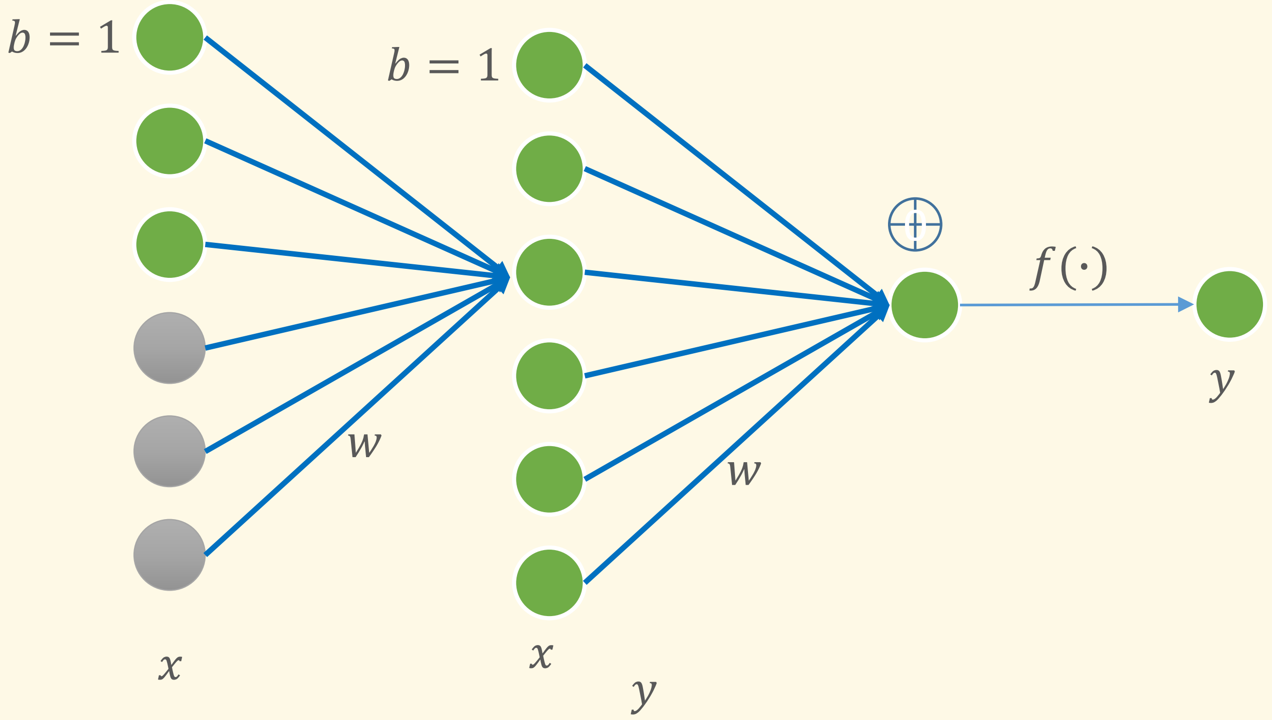
对于前馈神经网络

$$\vec{y} = f(\vec{x})$$

多层

$$\vec{y} = f_{total}(W \cdot \vec{x})$$

$$\vec{y} = f(W_1 \cdot f(W_2 \cdot \vec{x}))$$



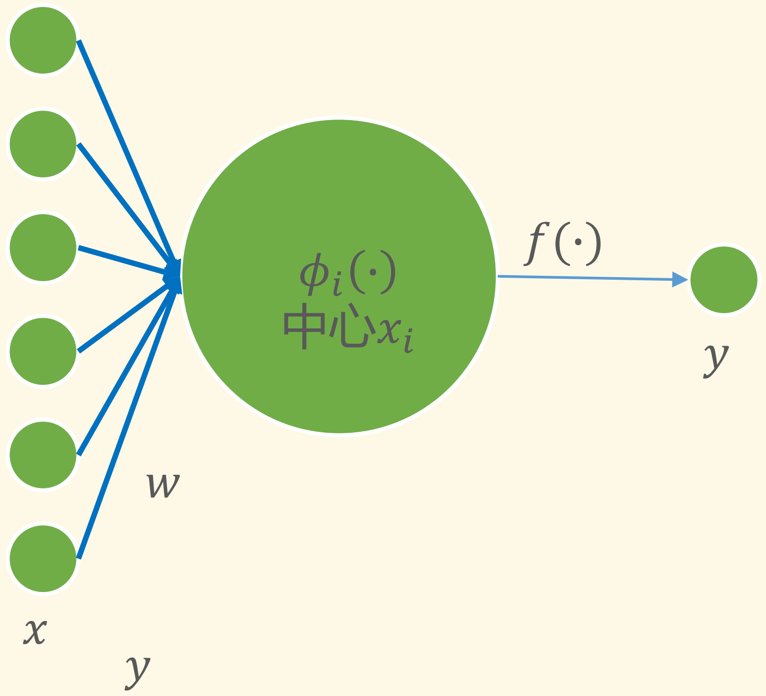
卷积神经网络

对于前馈神经网络

$$\vec{y} = f(\vec{x})$$

径向基函数网络

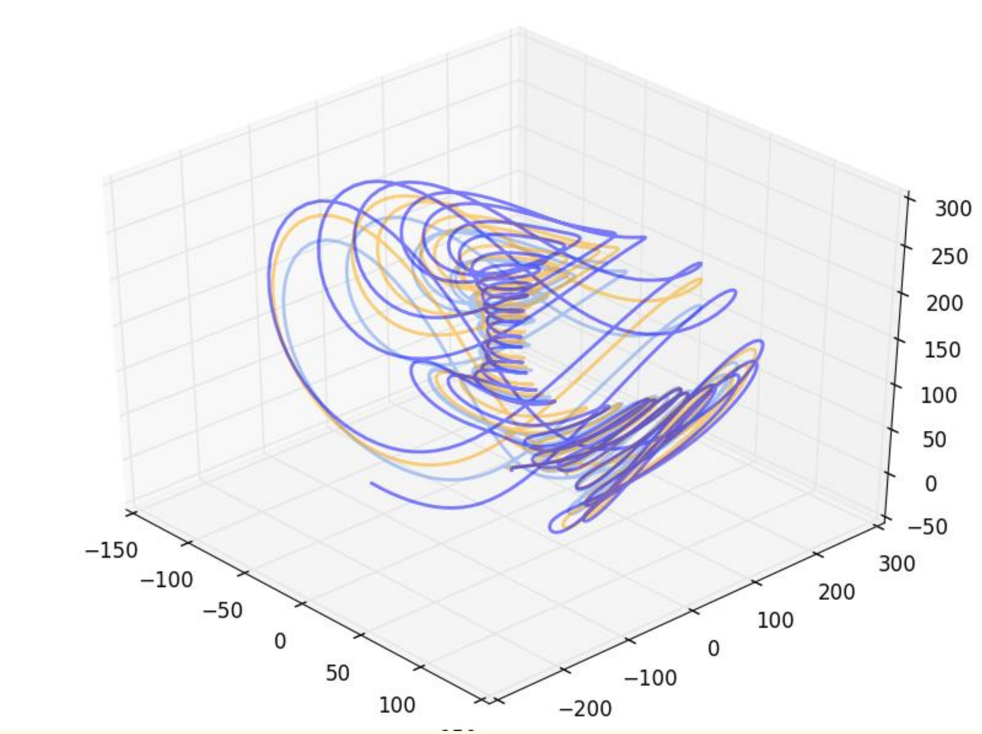
$$\vec{y} = f_{total}\left(W\phi(||\vec{x} - x_i||)\right)$$



对于动态神经网络

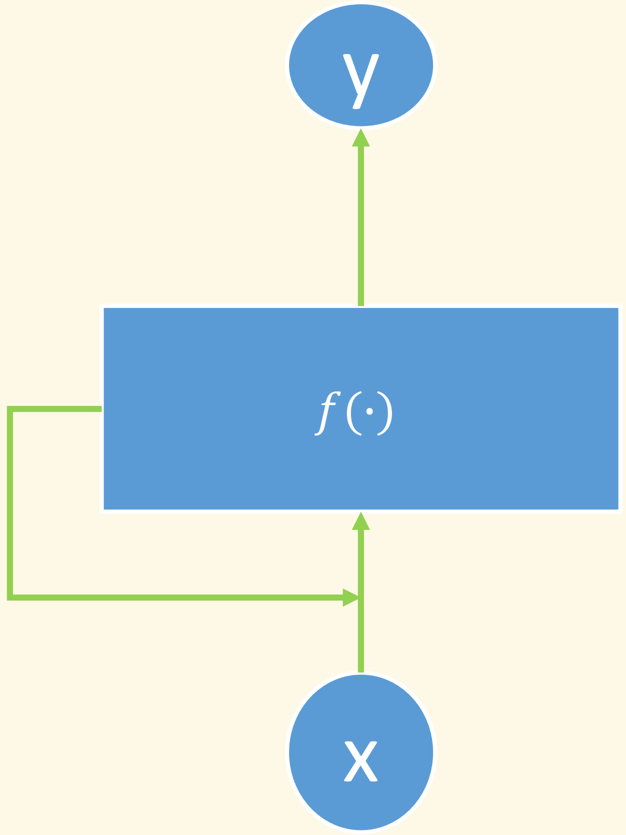
$$\vec{y} = f(\vec{x}; t)$$

直观理解
“混沌吸引子”



对于动态神经网络

$$\vec{y} = f(\vec{x}; t)$$



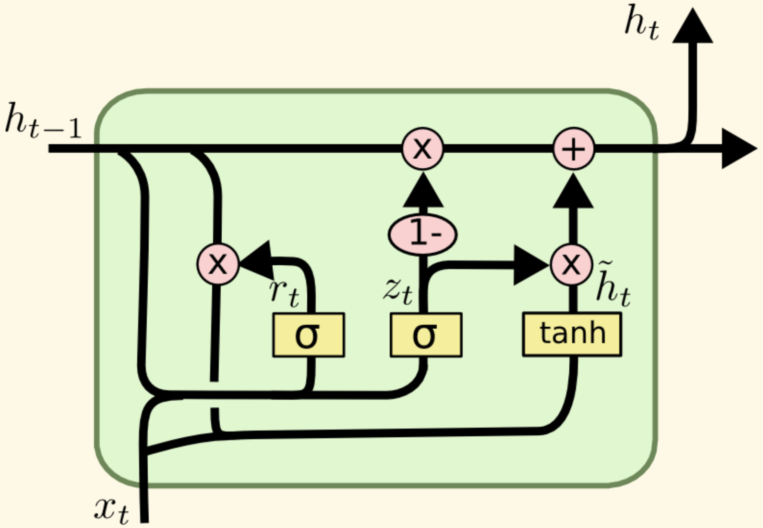
RNN

$$\vec{y}_{t+1} = f(W_1 \cdot \vec{x} + W_2 \cdot y_t)$$

对于动态神经网络

$$\vec{y} = f(\vec{x}; t)$$

LSTM



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

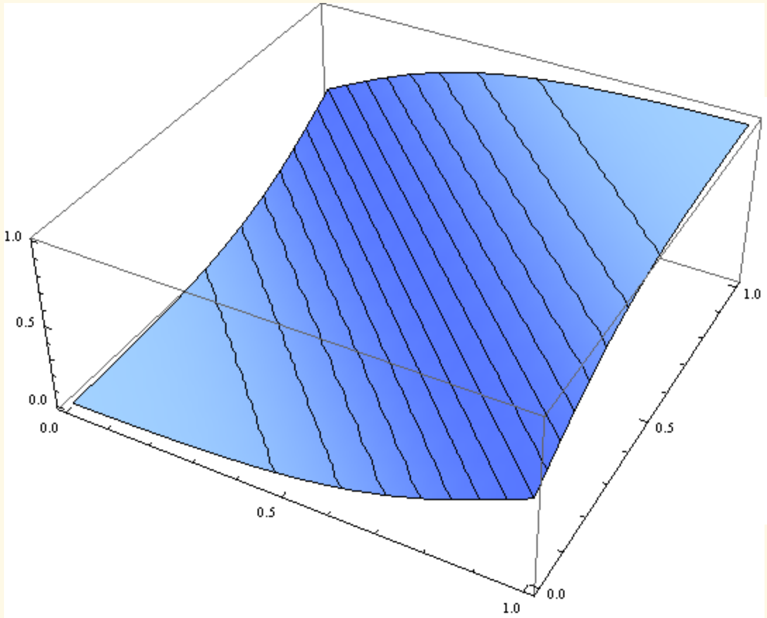
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

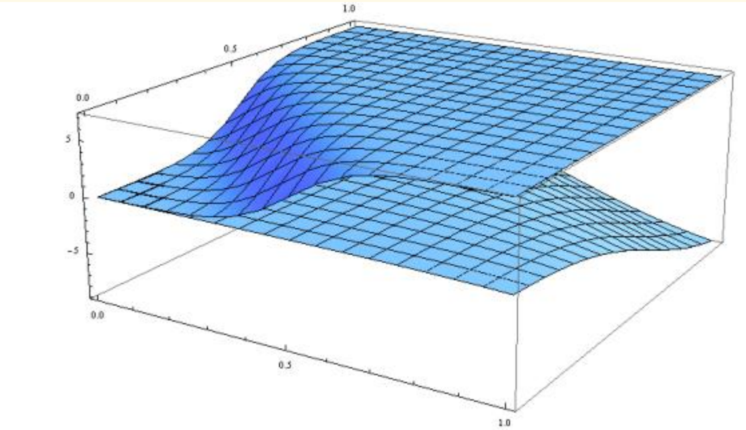
Part / 03

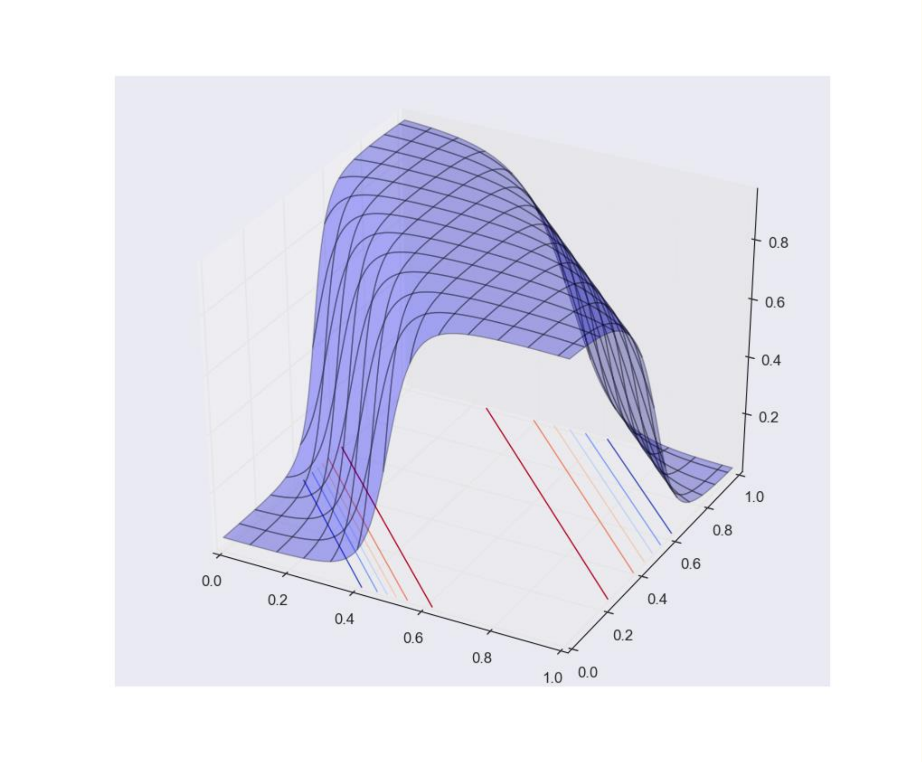
优化问题

OPMIZATION

经典抑或问题


1.0
0.5
0.0
0.0
0.5
1.0
0.0
0.5
1.0


1.0
0.5
0.0
5
0
-5
0.0
0.5
1.0


0.8
0.6
0.4
0.2
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0

优化问题

梯度问题

求解器

回顾函数展开 $f(\vec{x}) = f(\vec{x}_0) + \nabla f(\vec{x})\Big|_{x=x_0} \cdot d\vec{x} + d\vec{x}\underline{\underline{H}}\Big|_{x=x_0} d\vec{x}$

最速下降法 $f(\vec{x}) = f(\vec{x}_0) + \nabla f(\vec{x})\Big|_{x=x_0} \cdot d\vec{x} \Rightarrow d\vec{x} = -\eta\nabla f(\vec{x})$

牛顿法 $\mathrm{d}\left(\nabla f(\vec{x})\Big|_{x=x_0} \cdot d\vec{x} + d\vec{x}\underline{\underline{H}}\Big|_{x=x_0} d\vec{x}\right) = 0$

$d\vec{x} = -\underline{\underline{H}}^{-1}\nabla f(\vec{x})$(二阶连续可微)

求解器

BP算法

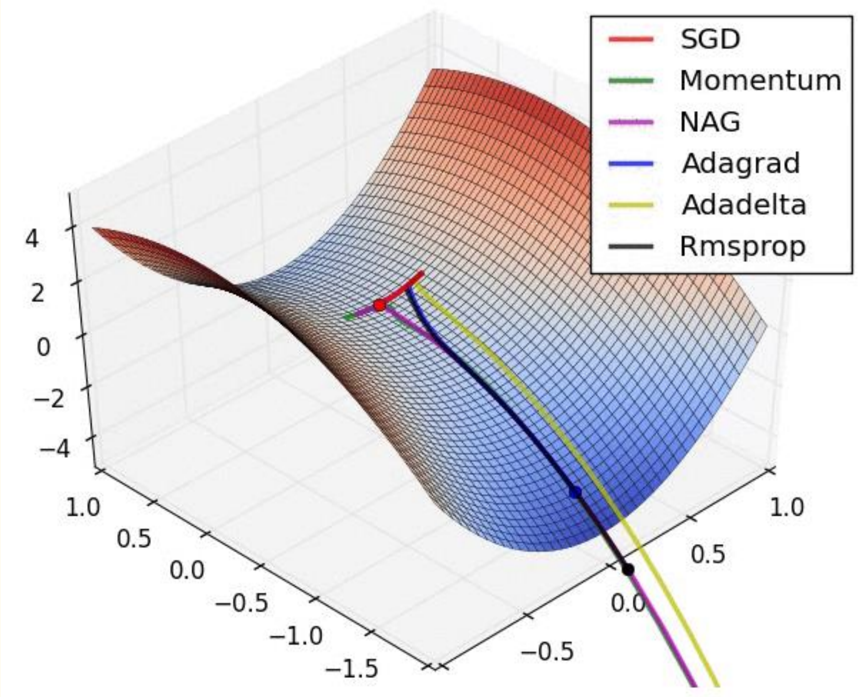
$$\nabla f(\vec{x}) = \frac{\partial f}{\partial w_{layer1}} + \frac{\partial f}{\partial w_{layer2}} + \frac{\partial f}{\partial w_{layer3}} + \ldots$$

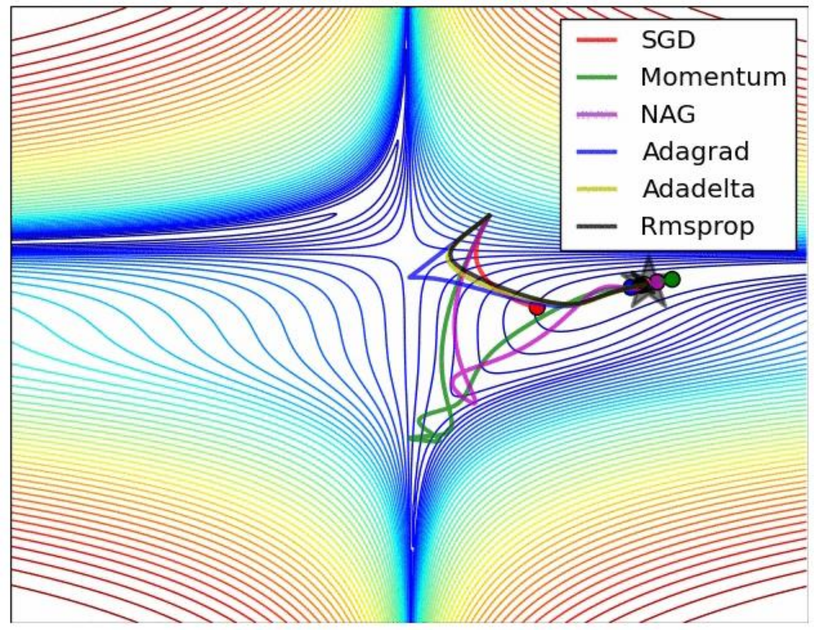
$$\frac{\partial f}{\partial w_{layer1}} = K \frac{\partial f}{\partial w_{layer2}}$$

Adaptive Subgradient

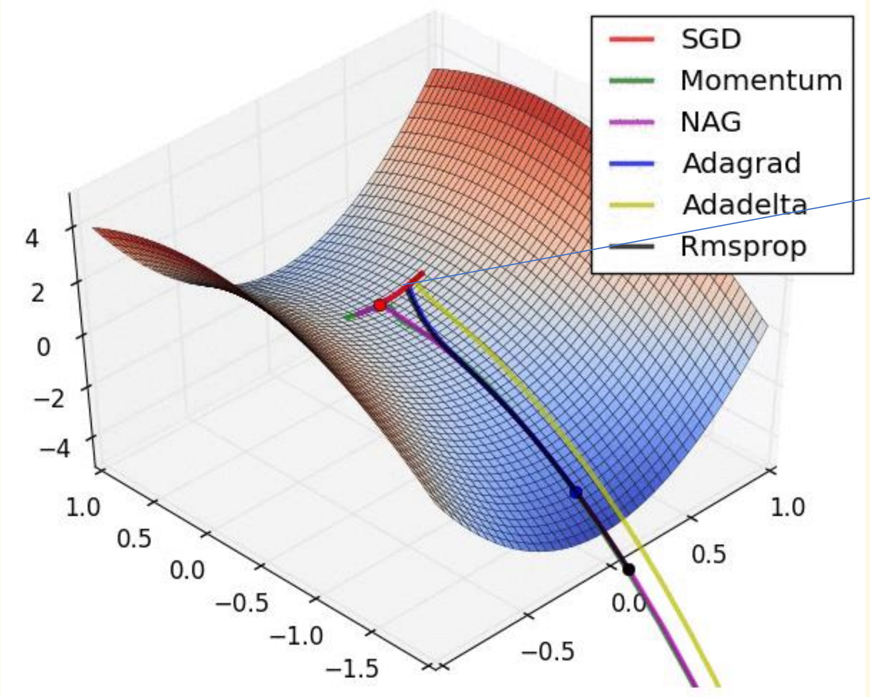
$$\frac{1}{\sqrt{\sum_{\tau=1}^{t-1}[\nabla f^{d}]}} \nabla f(\vec{x})$$

求解器直观理解


SGD
Momentum
NAG
Adagrad
Adadelta
Rmsprop
4
2
0
−2
−4
1.0
0.5
0.0
−0.5
−1.0
−1.5
1.0
0.5
0.0
−0.5

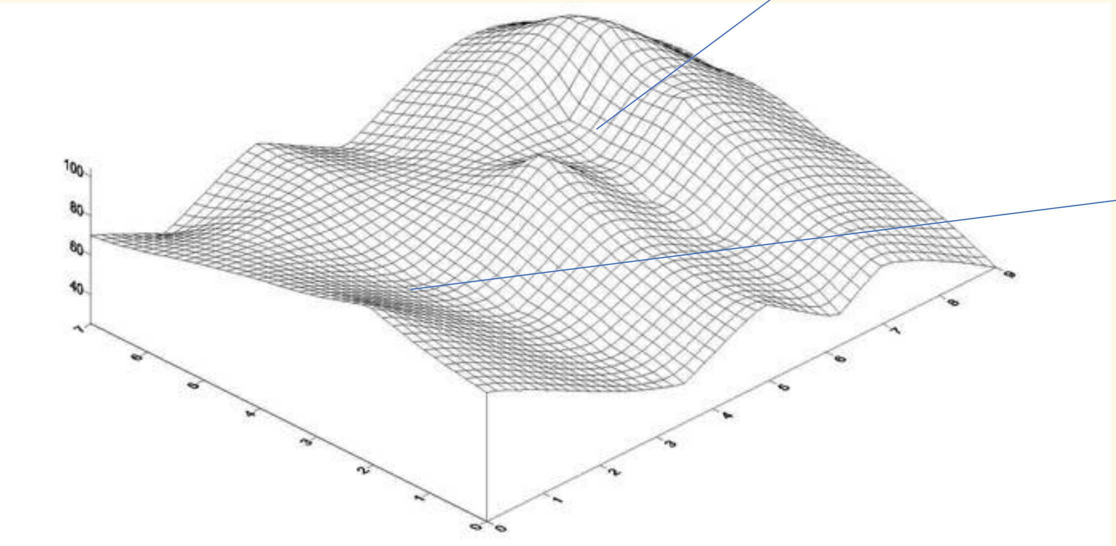

SGD
Momentum
NAG
Adagrad
Adadelta
Rmsprop

求解器问题



鞍点导致迭代收敛缓慢

求解器问题

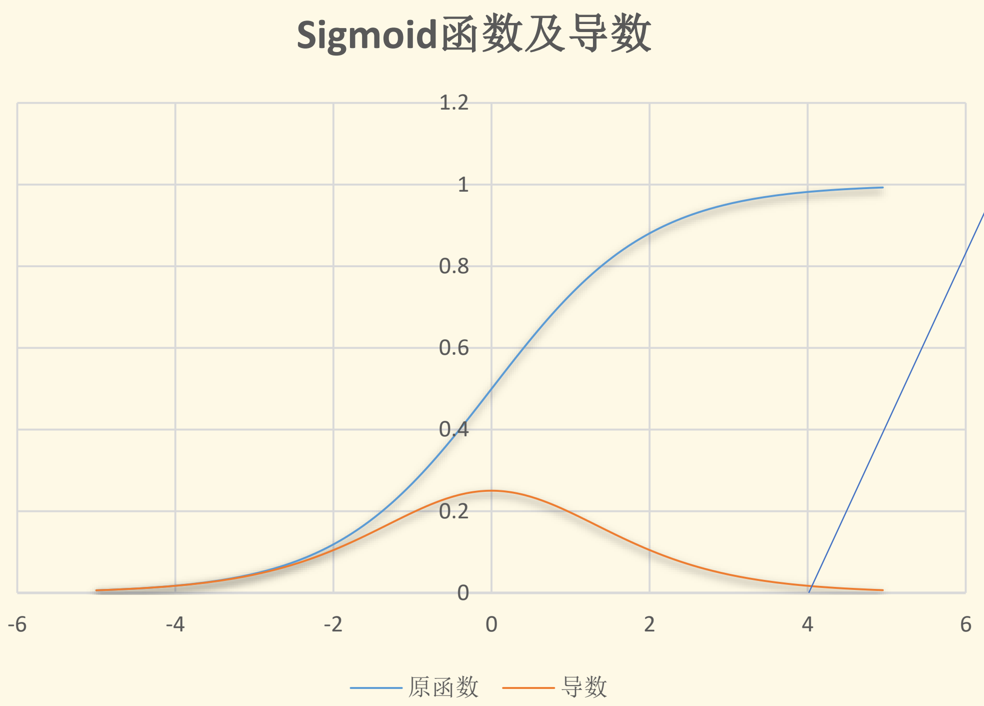


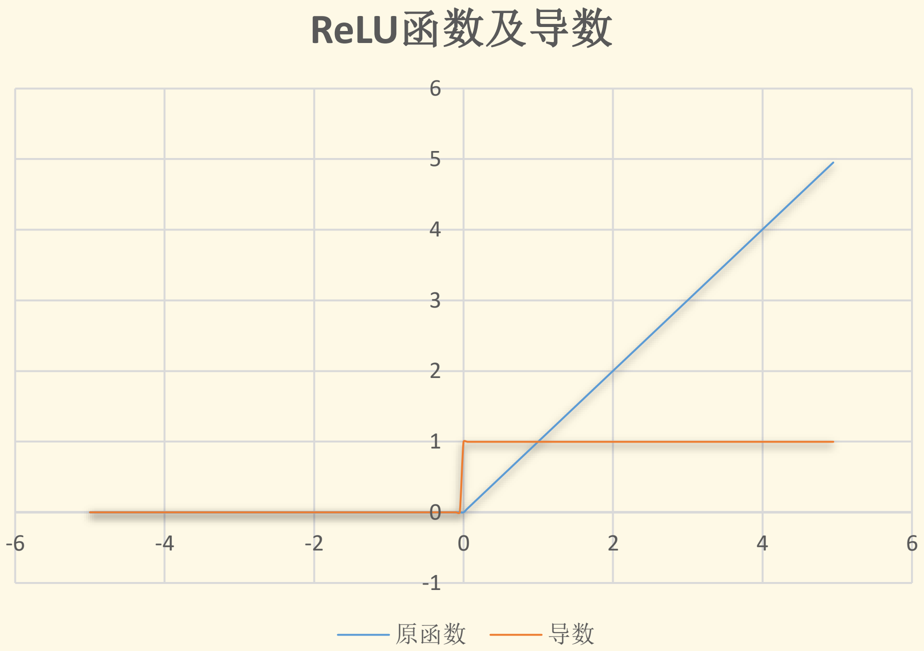
局部最小值

梯度消失

求解器问题

激活函数饱和引起梯度问题



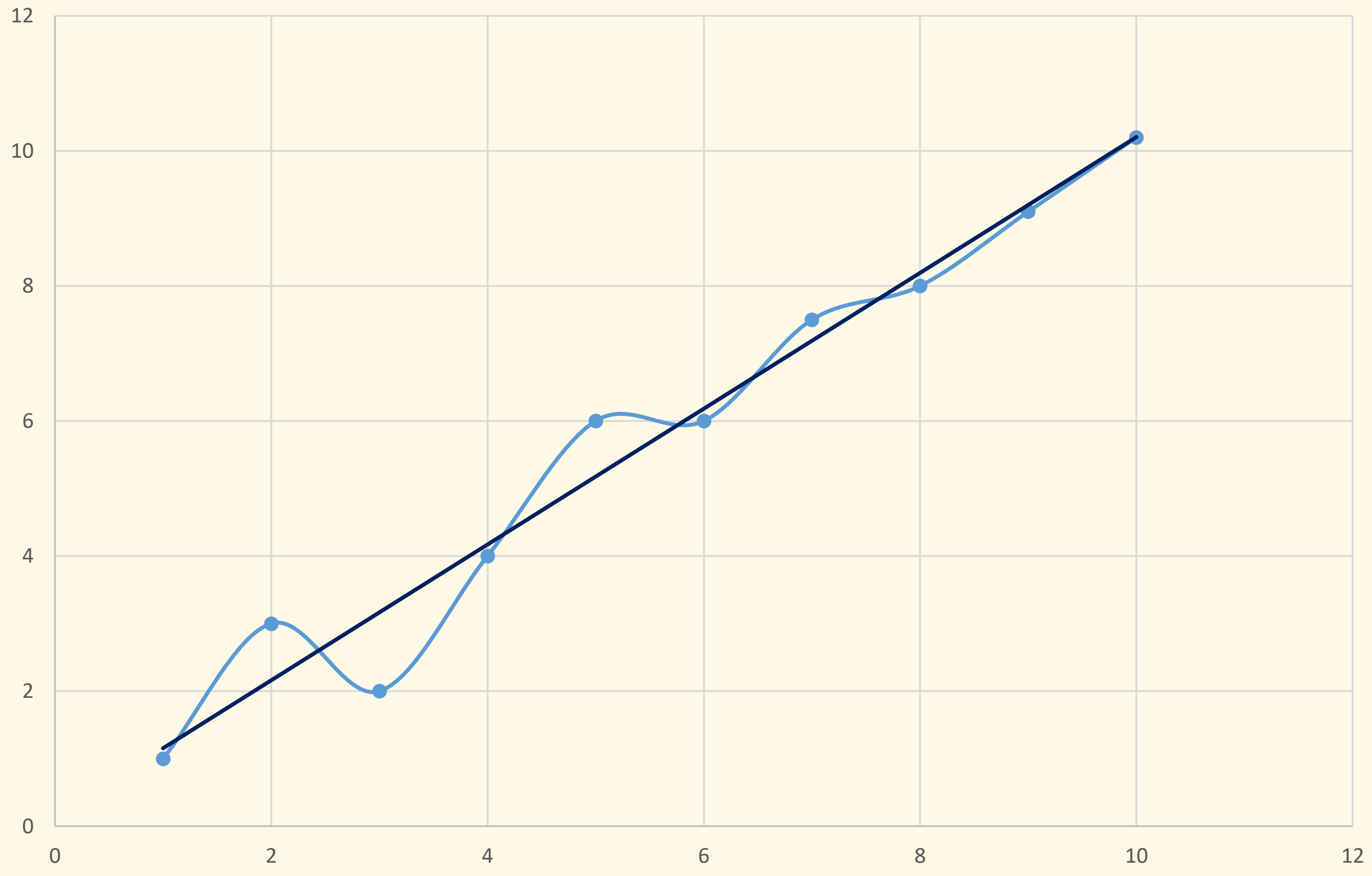


Part / 04

过拟合问题

OVERFITTING

过拟合问题



通过所有数据点的函数

过拟合问题

加入正则化项

$$\varepsilon = \varepsilon_s(f) + \frac{1}{2}||Df||^2$$

D线性微分算子

示例

$$\varepsilon = \varepsilon_s(f) + \frac{1}{2}\lambda||w||^2$$

D线性微分算子

过拟合问题

dropout

训练过程中随机选取一部分权值不进行训练

过拟合问题

Batch Nromalization 批归一化

Part / 05

应用实例

Example

神经网络应用

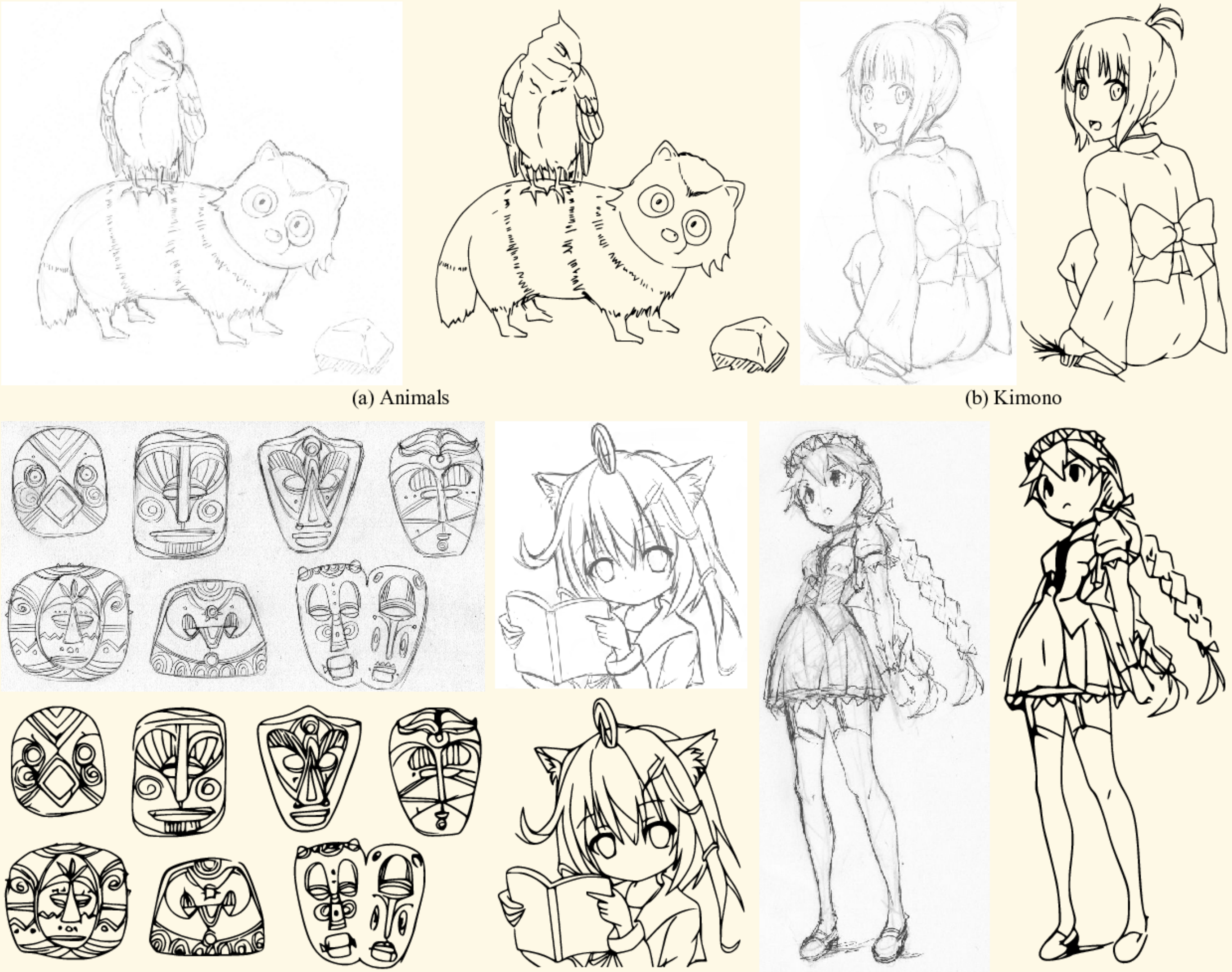
EXAMPLE

CNN

神经网络应用

EXAMPLE

CNN



(a) Animals (b) Kimono

(c) Masks (d) Book (e) Standing girl

神经网络应用

EXAMPLE

LSTM

THANKS

AI工程师讲座