

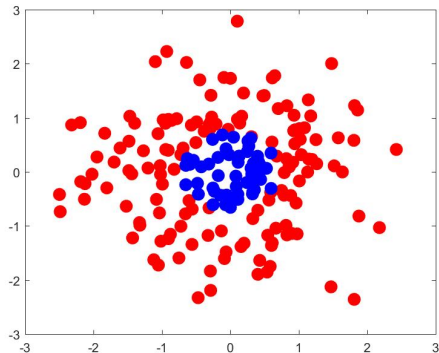
09 - Introduction to Nonlinear Models

Numerical Methods for Deep Learning

February 7, 2018

Why nonlinear problems

Impossible to find a linear separator between points



Goal/Trick

Embed the point in higher dimension or move the points to make them linearly separable

Why nonlinear problems

Consider regression try to find \mathbf{W} such that

$$\mathbf{C} = \mathbf{Y}\mathbf{W}$$

Assume $\mathbf{C} \in R^m$, $\mathbf{Y} \in R^{m \times n}$ and $m \gg n$.
If \mathbf{Y} is full rank impossible to fit the data.

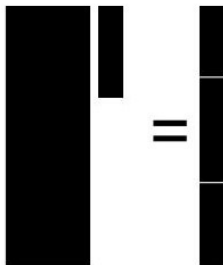
Why nonlinear problems

Consider regression try to find \mathbf{W} such that

$$\mathbf{C} = \mathbf{Y}\mathbf{W}$$

Assume $\mathbf{C} \in R^m$, $\mathbf{Y} \in R^{m \times n}$ and $m \gg n$.

If \mathbf{Y} is full rank over-determined impossible to fit the data.



Why nonlinear problems

Use a linear transformation of the points

$$\mathbf{C} = (\mathbf{Y}\mathbf{K})\hat{\mathbf{W}}$$

Assume $\mathbf{C} \in R^m$, $\mathbf{K} \in n \times s$ is an arbitrary matrix.
 $(\mathbf{Y}\mathbf{K}) \in R^{m \times s}$ and s arbitrary.

If \mathbf{Y} is full rank over-determined (still) impossible to fit the data.



A diagram illustrating a linear system of equations. It shows a matrix multiplication where a vertical black rectangle (representing a vector or matrix) is multiplied by a horizontal black rectangle (representing a vector or matrix). The result is shown as a single vertical black rectangle (representing a vector or matrix). The equation is represented as: $\begin{bmatrix} \blacksquare \\ \blacksquare \\ \blacksquare \end{bmatrix} \begin{bmatrix} \blacksquare & \blacksquare & \blacksquare \end{bmatrix} = \begin{bmatrix} \blacksquare \\ \blacksquare \\ \blacksquare \end{bmatrix}$

Why nonlinear problems



The size of \mathbf{YK} is $m \times s$ but the problem is still over-determined

$$\text{rank}(\mathbf{YK}) = n \ll m$$

Why nonlinear problems



A diagram illustrating the equation $\mathbf{YK} = \mathbf{b}$. On the left, a tall black rectangle represents matrix \mathbf{Y} (size $m \times n$), followed by a shorter, wider black rectangle representing matrix \mathbf{K} (size $n \times s$). A vertical line separates these from an equals sign, which is followed by a tall, thin black rectangle representing vector \mathbf{b} (size $m \times 1$).

$$\text{rank}(\mathbf{YK}) = n \ll m$$

But if we use a “smart” nonlinear function f then

$$\text{rank}(f(\mathbf{YK} + b)) = \min(m, s)$$

The fundamental theorem

Given the data $\mathbf{Y} \in R^{m \times n}$ and $\mathbf{C} \in R^m$ with $m > n$ There is a transformation f and a matrix $\mathbf{K} \in R^{n \times s}$ such that $f(\mathbf{YK})$ is of full rank m and therefore we can find a vector \mathbf{W} such that

$$f(\mathbf{YK} + b)\mathbf{W} = \mathbf{C}$$

The fundamental theorem

$$f(\mathbf{YK} + b)\mathbf{W} = \mathbf{C}$$

- ▶ How to find f ?
- ▶ How to find \mathbf{K} and b ?

The fundamental theorem

$$f(\mathbf{Y}\mathbf{K} + b)\mathbf{W} = \mathbf{C}$$

- ▶ How to find f ?
- ▶ How to find \mathbf{K} and b ?

Early days - motivated by neurons

- ▶ f is a smooth step function $\tanh(t)$
- ▶ Better choices for f is the relu $f = \max(x, 0)$
- ▶ \mathbf{K}, b can be
 - ▶ Random (random kernels recently branded as extreme learning machines)
 - ▶ Optimized (learned) to fit the data

Random kernels

Choose \mathbf{K} randomly and solve the least-squares/classification problem

- ▶ Can interpolate any function
- ▶ May require very large \mathbf{K} (size of the data)
- ▶ May not generalize well
- ▶ Large dense linear algebra
- ▶ Very easy to program.
- ▶ Can serve as a benchmark to more sophisticated methods

Random kernels - programming

Homework assignment I

- ▶ Regression: Given the data $\{\mathbf{Y}, \mathbf{C}\}$ write a code that computes the weights such that

$$f(\mathbf{YK})\mathbf{W} \approx \mathbf{C}$$

- ▶ Classification: Given the data $\{\mathbf{Y}, \mathbf{C}\}$ write a code that computes the weights such that $f(\mathbf{YK})\mathbf{W}$ classifies \mathbf{C}
- ▶ Try different activation functions and different size of \mathbf{K} .

Test your code on a validation set and report.

Prize to the best winning team