Setting optimization problem

- Normalize inputs
- Data vanishing and exploding gradients
- Weights initialization

Normalize inputs

- Z – score

$$X \coloneqq \frac{X - \mu}{\sigma}$$

- Objective: training deep neural network (L layers)
- $w^{[l]} = \begin{matrix} 1,5 & 0 \\ 0 & 1,5 \end{matrix} \rightarrow +\infty$: exploding gradient if L is big
- $w^{[l]} = \begin{matrix} 0,5 & 0 \\ 0 & 0,5 \end{matrix} \rightarrow 0$: Data vanishing if L is big
- Solution to this problem:
  Weight initialization

Weight initialization

- Set Variance of weights (of the same l):
  - $Var(w_i) = \frac{1}{n}$ with np,random,randn(shape) * np,sqrt(1/n[l-1]) of previous layer
  - If we use ReLU function $Var(w_i) = \frac{2}{n}$
  - Xavier initialization: for tanh activation function $\sqrt{\frac{1}{n^{[l-1]}}}$ or $\sqrt{\frac{2}{n^{[l-1]} * n^{[l]}}}$