

Gradient Descent for softmax regression

mapping: $f: \mathbf{x} \rightarrow \mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{b}$

where $\mathbf{x} \in \mathbb{R}^N$

$\mathbf{y}, \mathbf{b}, f(\cdot) \in \mathbb{R}^K$

$\mathbf{x} \in \mathbb{R}^N$

$\mathbf{W} \in \mathbb{R}^{K \times N}$

gradients: $\nabla_{W_{k,i}} \mathbf{L}_{CE} = x_i \left(\frac{\exp[\mathbf{w}_k \mathbf{x} + b_k]}{\sum_j \exp[\mathbf{w}_j \mathbf{x} + b_j]} - \delta_{1,y_k} \right)$

gradient update rule:

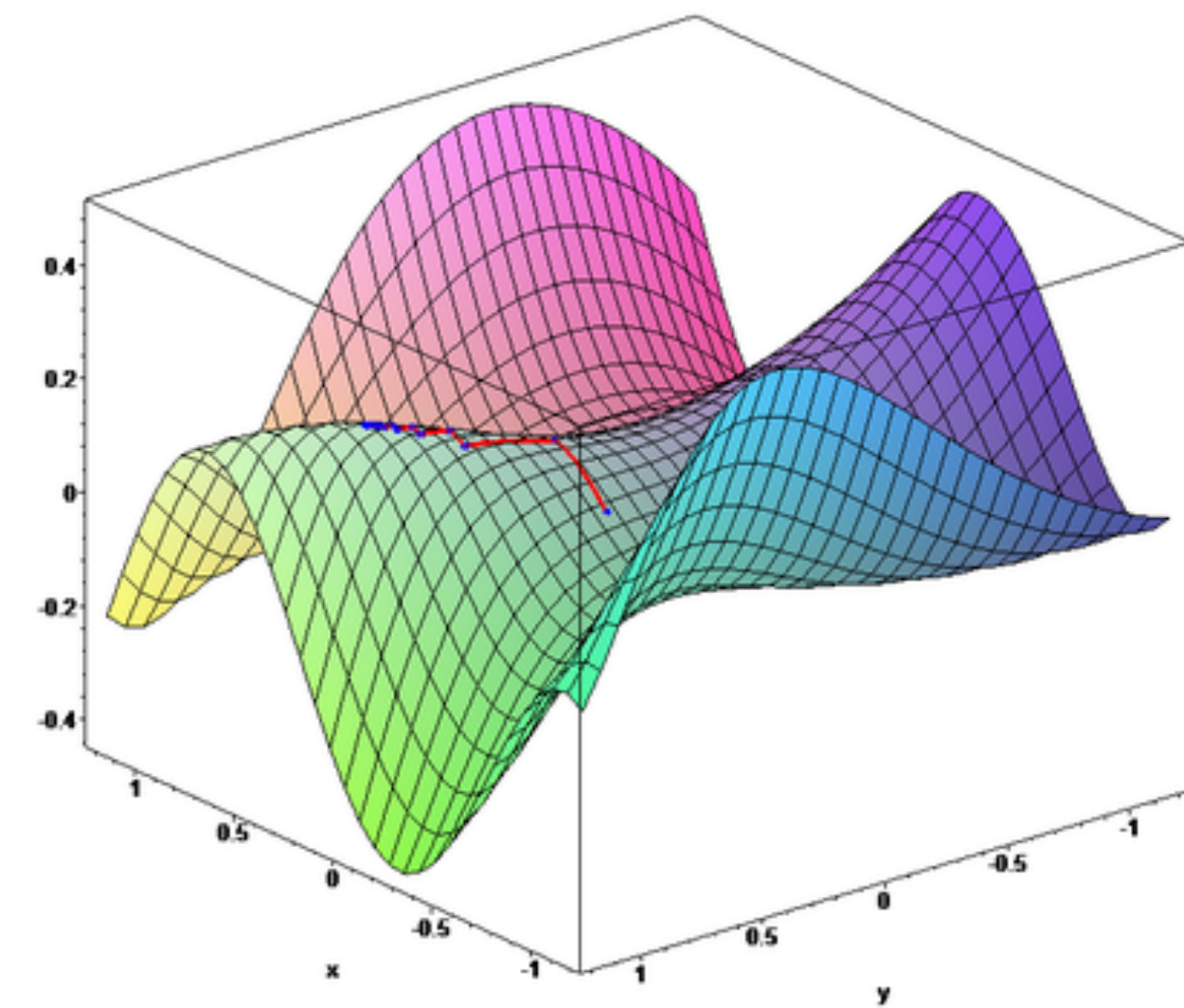
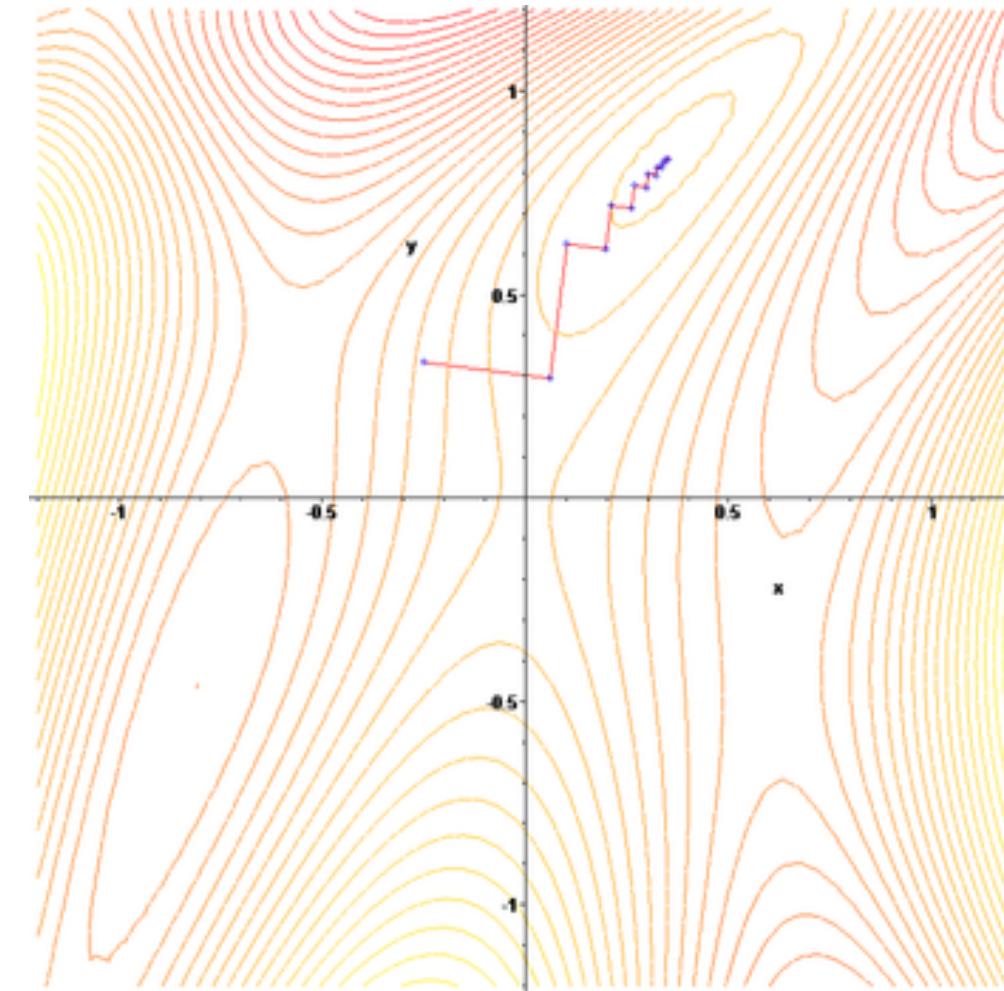
while not converged do : {

$\forall k, j$ do : {

$$W_{k,j} := W_{k,j} - \eta \nabla_{W_{k,j}} L_{CE}$$

}

}



- Source: Wikipedia



Georgetown
University

Lab 02: Text Normalization