2018 Turing Award: Deep Learning

S. Bengio [U Montreal]

G. Hinton [U Toronto]

Y. LeCun [NYU]

https://awards.acm.org/binaries/content/gallery/acm/awards/photo/a-b/bengio_3406375

https://awards.acm.org/binaries/content/gallery/acm/awards/photo/h-j/hinton_4791679

https://awards.acm.org/binaries/content/gallery/acm/awards/photo/k-l/lecun_6017366

Perceptron Learning Criterion

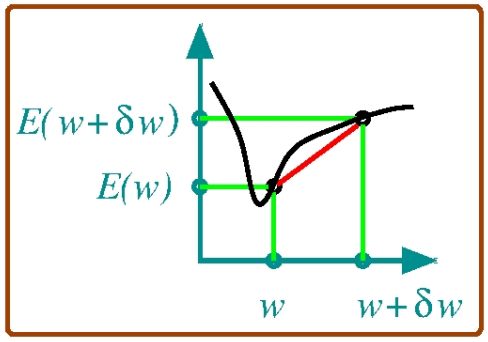
- 2-class classifier, i'th training point $\mathbf{x}^i$: $t_i = \pm 1$
- $y(\mathbf{x}^i) = h(\mathbf{w}^T\mathbf{x}^i) = h(a) = +1(a \geq 0), = -1(a < 0)$
- Declare Class $\mathscr{C}_1(t_i = +1)$ if $\mathbf{w}^T\mathbf{x}^i \geq 0$
- Declare Class $\mathscr{C}_2(t_i = -1)$ if $\mathbf{w}^T\mathbf{x}^i < 0$
- Combined: (SVM-like!) Correct, if $t_i\ \mathbf{w}^T\mathbf{x}^i \geq 0$
- Perceptron penalty: 0 if correct , else sum
- Simple criterion: $E(\mathbf{w}) \stackrel{\triangle}{=} -\sum_{\forall i \in \mathscr{M}} t_i\ \mathbf{w}^T\mathbf{x}^i$
- no penalty for 'how much' misclassification
- Solved numerically, Cauchy's weight update rule

Iterative Weight Update: Learning

https://upload.wikimedia.org/wikipedia/commons/d/d3/Augustin-Louis_Cauchy_1901.jpg

- A.-L. Cauchy [1789-1857]
- **Cauchy's Rule[1849]**
- $\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} - \eta \nabla E(\mathbf{w})$

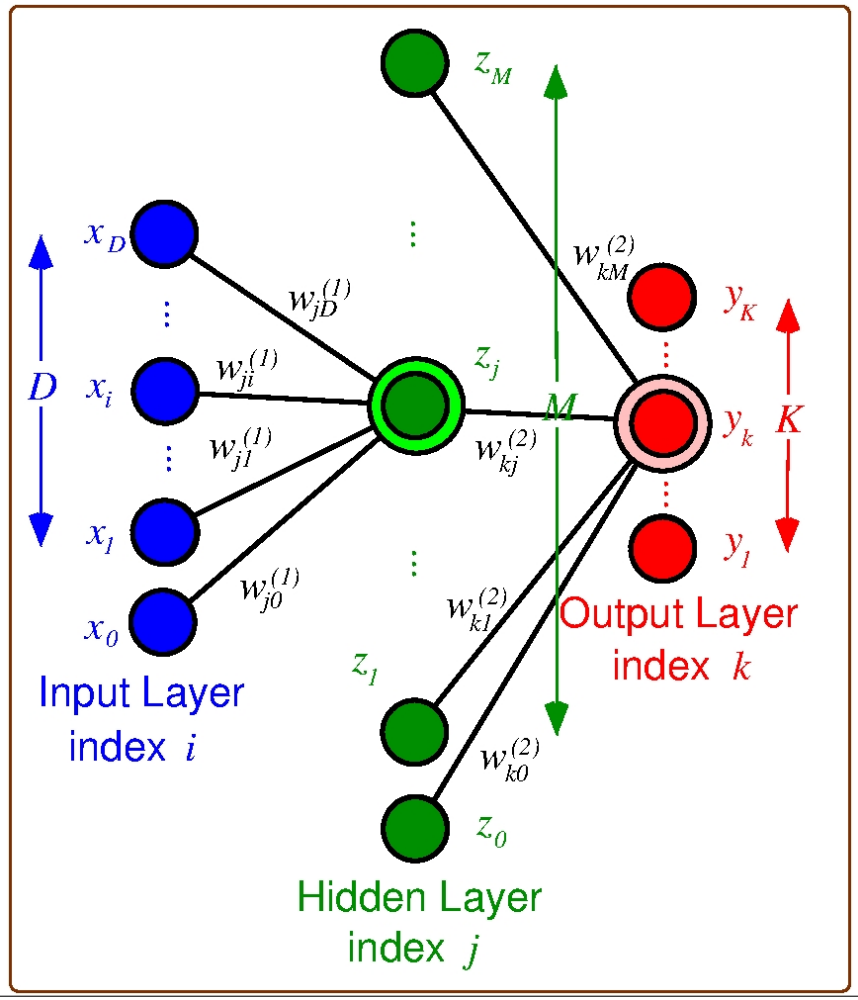
- dim cons?
- Step 'η'?
- why '-'?



- sl = $\frac{E(w+\delta w)-E(w)}{(w+\delta w)-(w)} = \frac{E(w+\delta w)-E(w)}{\delta w}$
- $\lim_{\delta w \to 0} : \partial E/\partial w \Longrightarrow \nabla E(\mathbf{w})$
- $\mathbf{w}$: $[\ w_j\ ]$, $\nabla E(\mathbf{w}) = [\ \partial E/\partial w_j\ ]$
- '-' sign: go against the gradient!

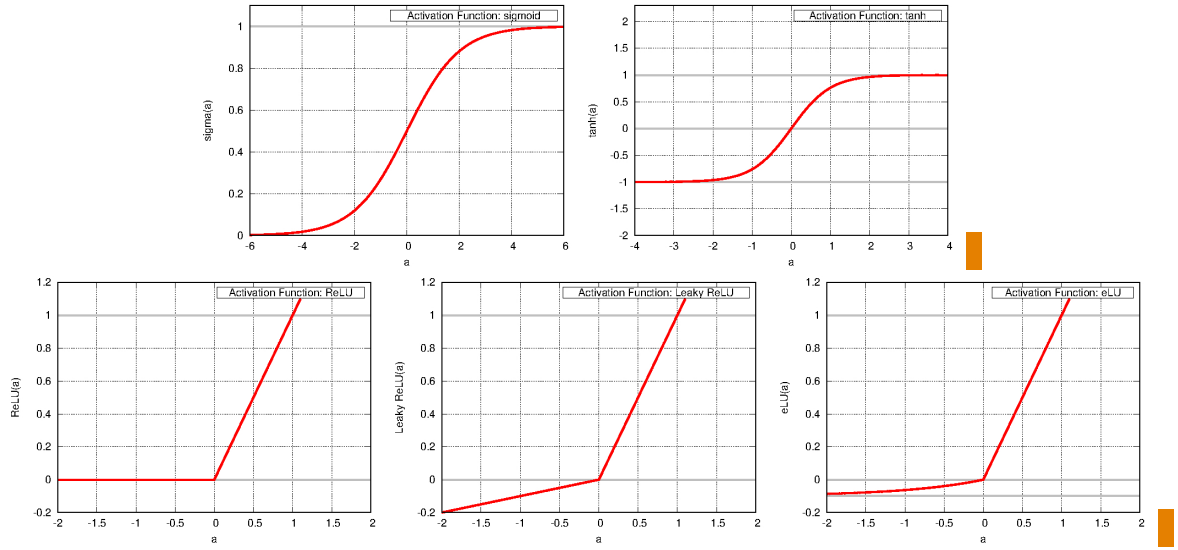
- 'η': step size or learning rate, tuning (adaptive?)
- small η: small steps, longer attain local min
- large η: large steps, may miss local min

Multi-Layer Perceptron



Activation Functions

- Neuron input: scalar, sum of weighted inputs
- Activation fn: possible non-linearity, scalar output
- X'fer fns: sigmoid, tanh, RelU, Leaky ReLU, eLU



- $LeakyRelU(a,\alpha) \triangleq \max(\alpha a, a),\ \alpha \in (0,1)$
- $eLU(a,\alpha) \triangleq \left\{ \begin{array}{c} a, a > 0) \\ \alpha(e^a - 1), a \leq 0 \end{array} \right\}$ [00:50]