VC-dimension of neural networks

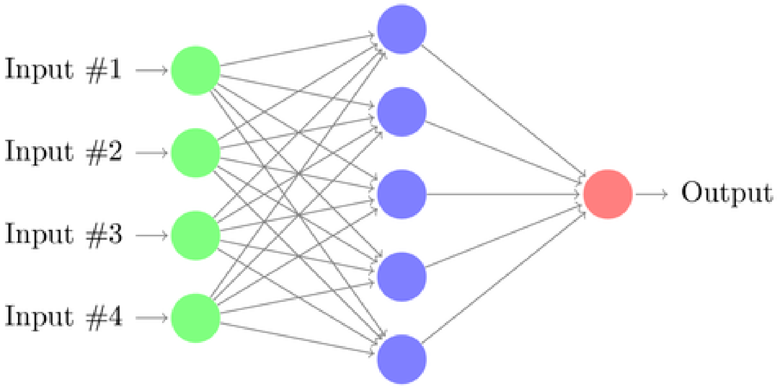
Abbas Mehrabian

Simons Research Fellow

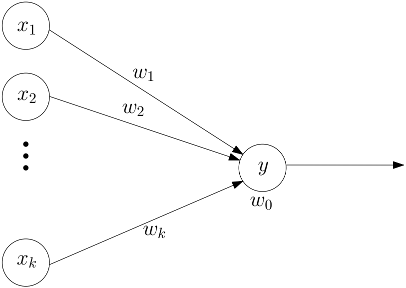
Simons Institute Industry Day
23 March 2017

joint with Nick Harvey and Chris liaw
University of British Columbia

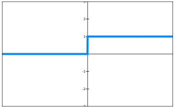
Neural networks



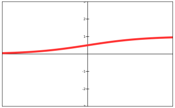
Neural networks



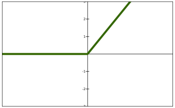
$$y = \sigma(w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_k x_k)$$



threshold



sigmoid



ReLU

How many data samples is needed to learn?

How many data samples is needed to learn?

Theorem (fundamental theorem of statistical learning Blumer, Ehrenfeucht, Haussler, Warmuth'89)

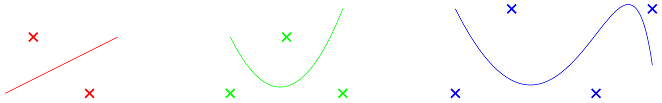
The number of samples required for PAC learning within error ε in a model with VC-dimension v is essentially $\Theta(v/\varepsilon)$.

How many data samples is needed to learn?

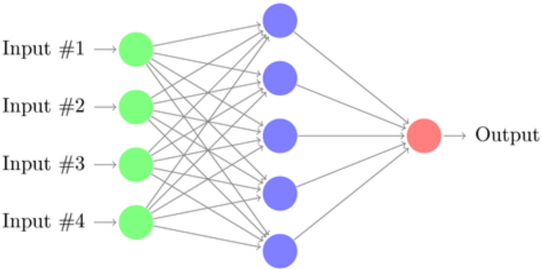
Theorem (fundamental theorem of statistical learning Blumer, Ehrenfeucht, Haussler, Warmuth'89)

The number of samples required for PAC learning within error ε in a model with VC-dimension v is essentially $\Theta(v/\varepsilon)$.

e.g. VC-dimension of polynomials of degree d is $d+1$.



VC-dimension of neural networks



$v(e, \ell) :=$ maximum VC-dimension of a neural network

e edges

ℓ layers

VC-dimension of neural networks

If the activation function is piecewise polynomial,

$$v(e,\ell) \le Ce^2 \qquad \text{[Goldberg, Jerrum'95]}$$

$$ce\ell \le v(e,\ell) \le C(e\ell^2 + e\ell\log e) \qquad \text{[Bartlett, Maiorov, Meir'98]}$$

VC-dimension of neural networks

If the activation function is piecewise polynomial,

$$v(e,\ell) \le Ce^2 \qquad \text{[Goldberg, Jerrum'95]}$$

$$ce\ell \le v(e,\ell) \le C(e\ell^2 + e\ell\log e) \qquad \text{[Bartlett, Maiorov, Meir'98]}$$

Theorem (Harvey, Liaw, M'16)

If the activation function is piecewise linear,

$$ce\ell\log(e/\ell) \le v(e,\ell) \le Ce\ell\log e$$

VC-dimension of neural networks

If the activation function is piecewise polynomial,

$v(e,\ell) \le Ce^2$ [Goldberg, Jerrum'95]

$ce\ell \le v(e,\ell) \le C(e\ell^2 + e\ell \log e)$ [Bartlett, Maiorov, Meir'98]

Theorem (Harvey, Liaw, M'16)

If the activation function is piecewise linear,

$$ce\ell \log(e/\ell) \le v(e,\ell) \le Ce\ell \log e$$

GoogleNet'14: $\ell = 41$, $e = 7$ million