Announcements

HW 4 Sarsa λ, Q-learning
Project Ideas
Project: Group of up to 3
4 page

HW 3

Reinforcement Learning

Model Based $\longleftrightarrow$ Model Free

Learn TR
solve

On Policy — Off Policy

Learn π: Policy Grad

[Learn Q: Sarsa(λ) — Q-learning]

Eligibility Traces — Double Q

Challenges
1. Exploration / Exploitation : Bandits
2. Credit Assignment
3. Generalization

So far, Tabular
Now: Neural Network Function Approx.

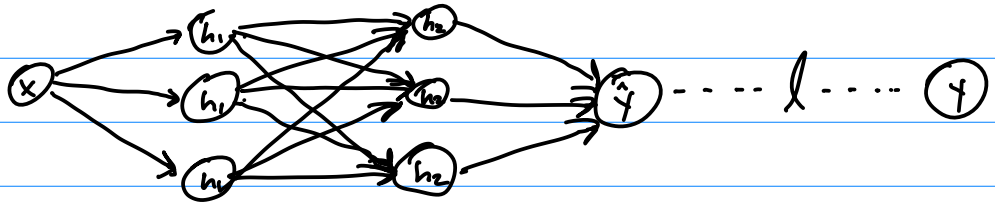
Linear $f_\theta(x) = \theta^T \beta(x)$

$\hat{y} = f_\theta(x)$ (input)

$\hat{y}$: output; θ: parameters

(x, y)

$\ell(\hat{y}, y) = (\hat{y} - y)^2$

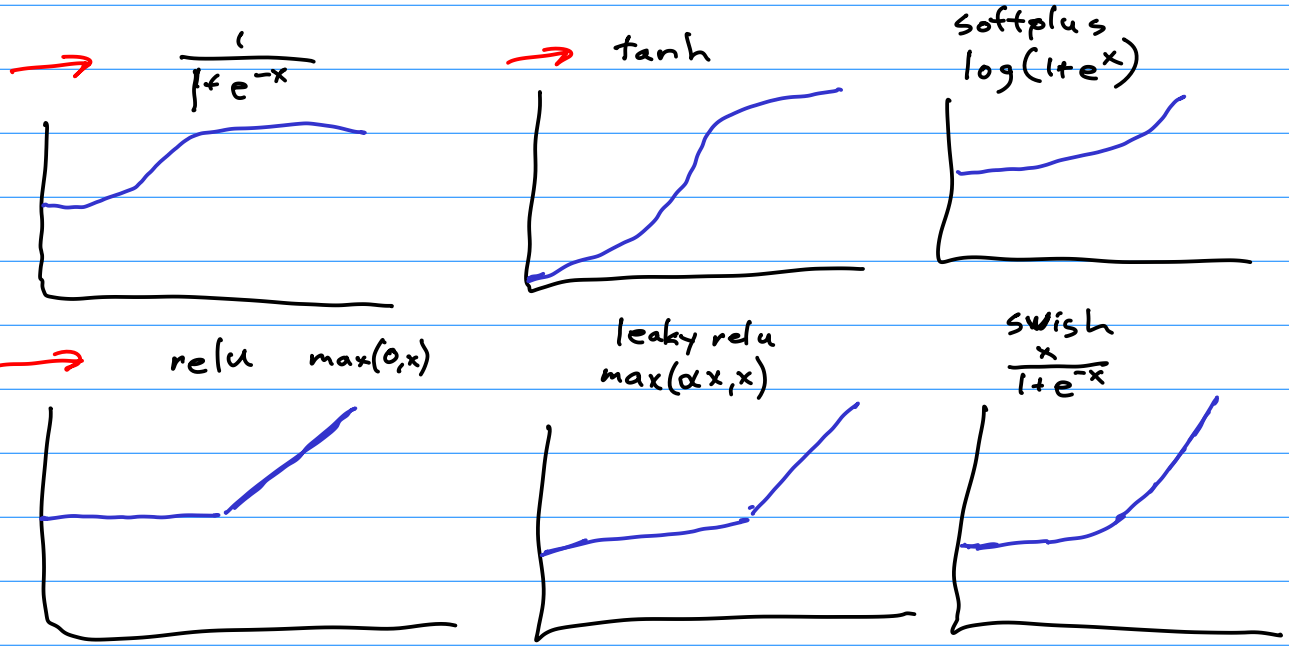


$h_i(x) = \sigma_i(w_i^T x + b_i)$

σ_i: nonlinearity; w_i: weights; b_i: bias

$$\hat{y} = W_3\, \sigma_2\left(W_2\, \sigma_1\left(W_1 x + b_1\right) + b_2\right) + b_3$$

σ



Training

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sum_{(x,y)\in D} l\left(f_\theta(x), y\right) \qquad \hat{y} = f_\theta(x)$$

Stochastic Gradient Descent

$$\theta \leftarrow \theta - \alpha \nabla_\theta\, l\left(f_\theta(x), y\right)$$

↖ software will do this for you

$f(g(h(x)))$ Chain Rule

$$\frac{\partial f(g(h(x)))}{\partial x} = \frac{\partial f(g(h))}{\partial h}\,\frac{\partial h(x)}{\partial x} = \frac{\partial f(g)}{\partial g}\,\frac{\partial g(h)}{\partial h}\,\frac{\partial h(x)}{\partial x}$$

↑

$$\hat{y} = W_2\, \sigma\left(W_1 x + b\right) + b_2$$

Backpropagation

$$\frac{\partial l}{\partial W_2} = \frac{\partial l}{\partial \hat{y}}\left(\frac{\partial \hat{y}}{\partial W_2}\right)^T$$

$$= \frac{\partial l}{\partial \hat{y}}\left(\sigma\left(W_1 x + b_1\right)\right)^T$$

The Alignment Problem

Conv Net

$$\underset{\theta}{\operatorname{argmin}} \sum_{(x,y)\in D} \ell(f_\theta(x), y) - \beta \|\theta\|^2$$

Dropout

Deep → More Layers