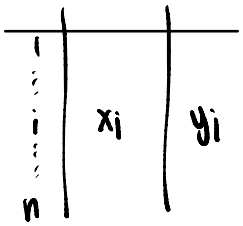
LECTURE 19

review so far

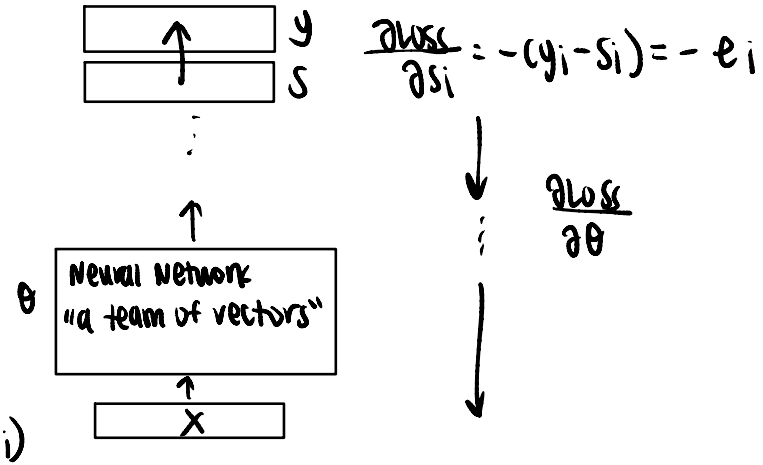
supervised:



learn $P_\theta(y|x)$

max log likelihood

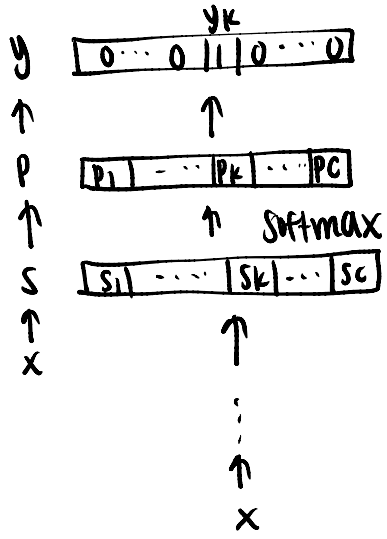
$$\text{log-lik}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \log P_\theta(y_i|x_i)$$



continuous y (regression)

$$\text{-log-likelihood} = \text{loss} = \frac{1}{n}\sum_{i=1}^{n} (y_i - s_i)^2/2$$

categorical y (classification)



category / word

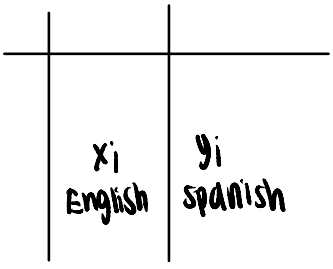
$$\text{log-likelihood} = \frac{1}{n}\sum_{i=1}^{n} \left\langle \begin{bmatrix} \\ \\ \end{bmatrix}_{y_i}, \begin{bmatrix} \\ \\ \end{bmatrix}_{s_i} \right\rangle - \log z$$

$$z = \sum_{k=1}^{C} e^{s_k}$$

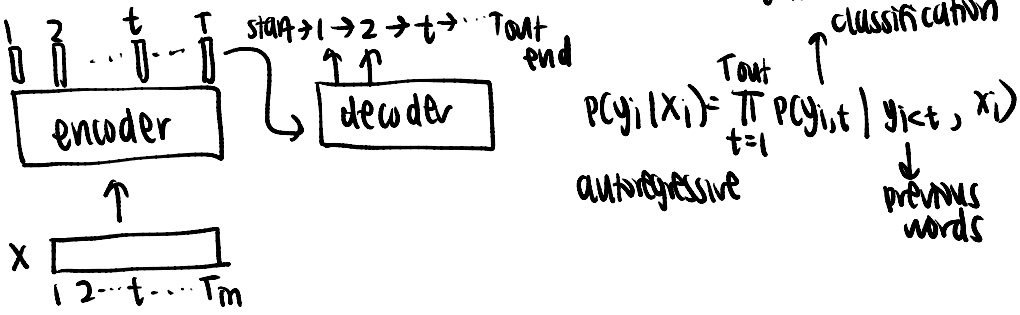
$$\text{loss} = -\text{log-likelihood}$$

$$\frac{\partial loss}{\partial s_i} = -\left(\begin{bmatrix} \\ \\ \end{bmatrix}_{y_i} - \begin{bmatrix} \\ \\ \end{bmatrix}_{p_i} \right) = -e_i$$

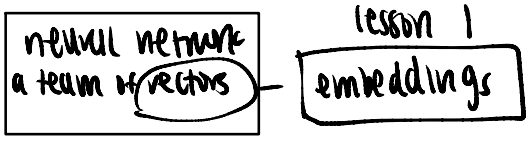
translation

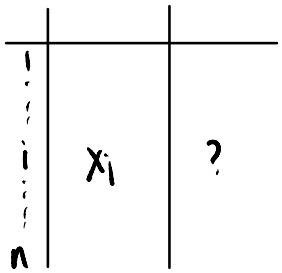


$$\text{log-lik}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \log P(\underbrace{y_i}_{\text{spanish}}|\underbrace{x_i}_{\text{English}})$$

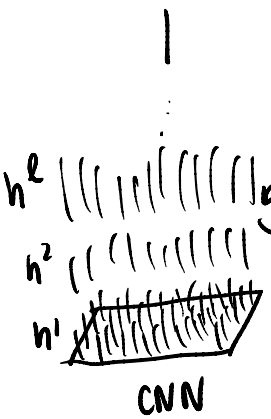


unsupervised:

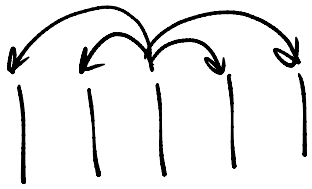


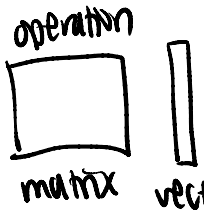


$$\frac{1}{n}\sum_{i=1}^{n} \log P_\theta(x_i)$$



$$h^{l,t+1} = h^{l,t} + f^{l,t}(h^{l,t})$$





like a computer program

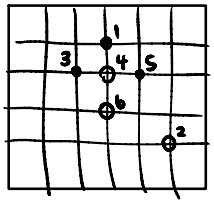
lesson 2

learned computation → multiple steps

instead of multiple layers

reinforcement learning (RL)

Alpha Go

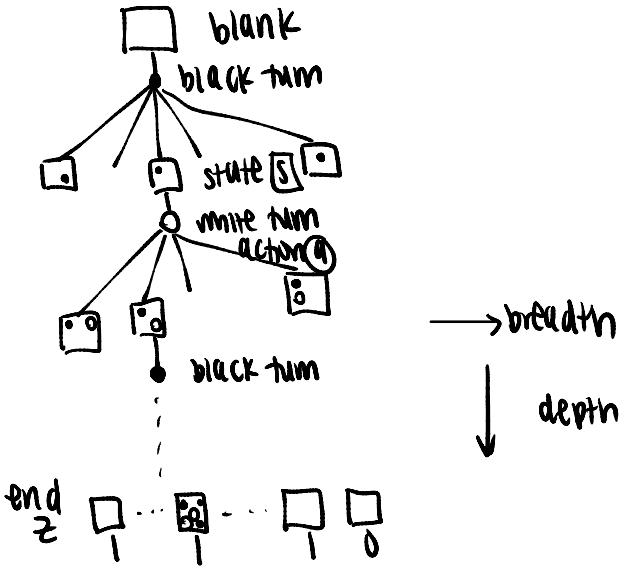


● black
○ white

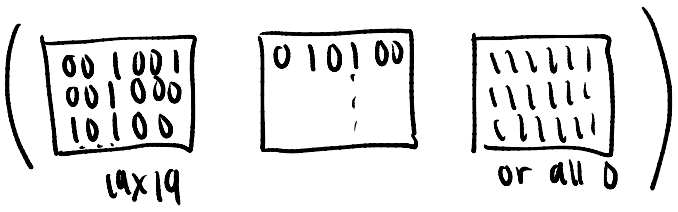
$$z = \begin{cases} 1 & \text{if } \bullet \text{ wins} \\ -1 & \text{if } \circ \text{ wins} \\ 0 & \text{if draw} \end{cases}$$

19×19

Game tree

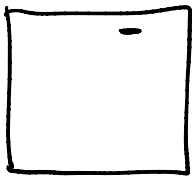


state S:



1: black stone 1: white stone all 1: black turn all 0: white turn

action a:



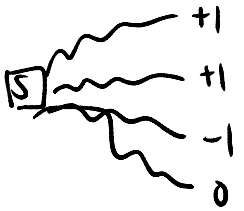
(19x19 + 1) categories
↓
pass

① policy: $P(a|S)$ classification

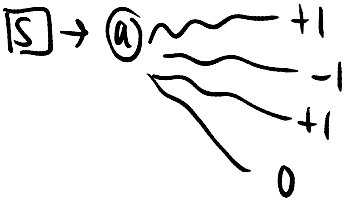
② value (state): $V(S) = E_{policy}(z|S)$ (e.g. .7)

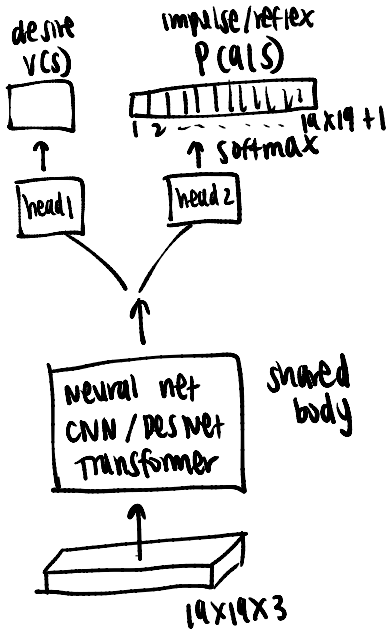
$S \rightarrow a \rightarrow \cdots \rightarrow \square \rightarrow z$

average



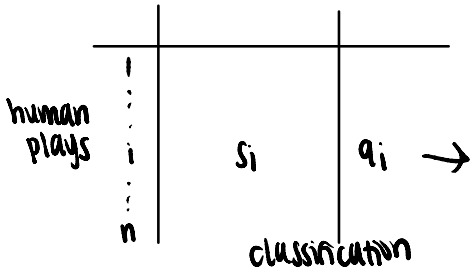
② value (state, action): $Q(S,a) = E_{policy}(z|S,a)$





learning:

Step 1: supervised learning



$\rightarrow P_\sigma(a|s)$ ↓ supervised

$$\max_\sigma \frac{1}{n}\sum_{i=1}^{n} \log P_\sigma(a_i|s_i)$$

Imitation, behavior cloning
Stochastic gradient ascent

$$\Delta\sigma \propto \frac{\partial}{\partial\sigma}\log P_\sigma(a_i|s_i)$$

① generated by human

Step 2: RL

$P_\rho(a|s)$

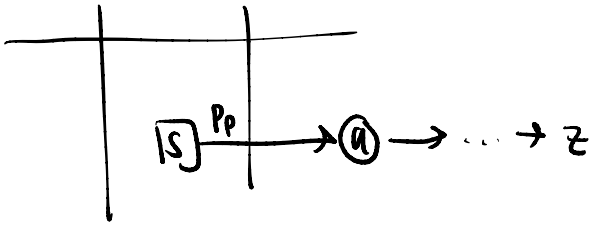
reinforcement

each follows own policy

initialize $\rho = \sigma$

let ρ play against σ

repeat



Policy Gradient (how we update the policy)

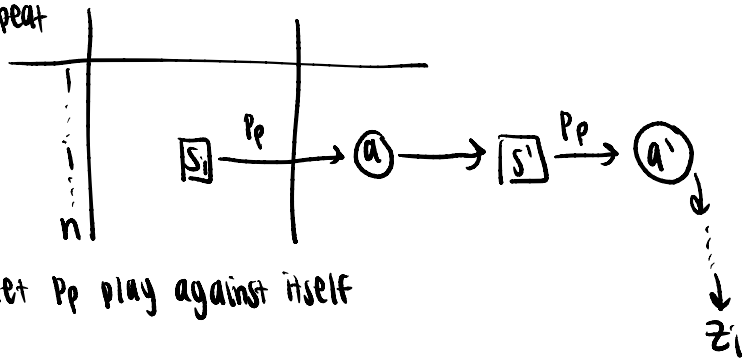
$$\Delta\rho \propto \frac{\partial}{\partial\rho}\log P_\rho(a|s)\cdot z$$

② reinforcement
$z = +1$ → good action, ↑ prob. of action
$z = -1$ → bad action ↳ want ↓ prob. of action
$z = 0$

① generated by P_ρ self try out different actions & see what will happen

Step 3: value network

repeat



let p_ρ play against itself

regression

$$\min_\theta \frac{1}{n} \sum_{i=1}^{n} (\overbrace{z_i}^{y} - v_\theta(\overbrace{s_i}^{x}))^2$$

$$\Delta\theta \propto -\frac{\partial}{\partial\theta}(z - v_\theta(s))^2$$