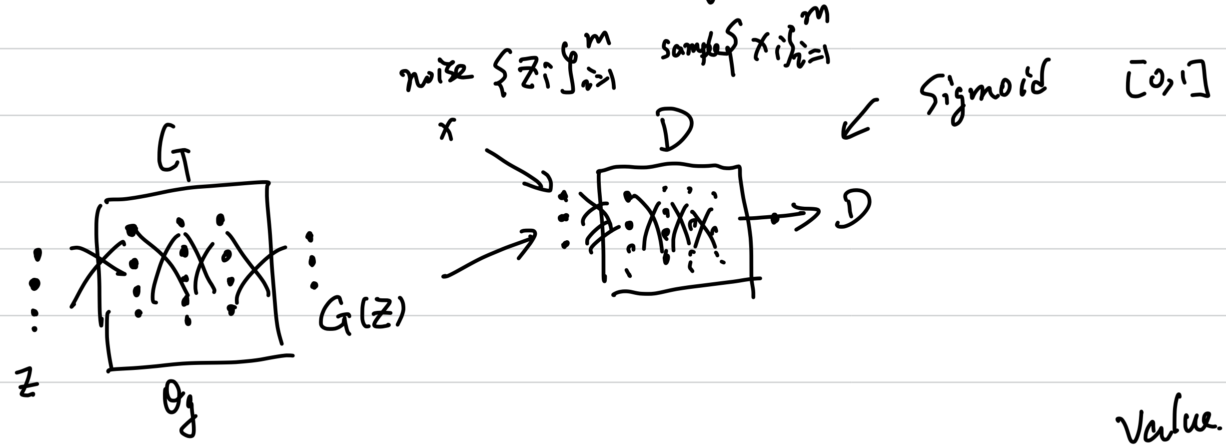
机器学习第六讲



Value.

$$V(D,G)=\frac{1}{m}\sum_{i=1}^{m}\log D(x_i)+\frac{1}{m}\sum_{i=1}^{m}\log\left(1-D(z_i)\right)$$

① 训练 D， $V_G=\max_{D} V(D,G)$ fixed G

$$\theta_d^{t+1}=\theta_d^{t}+\eta_d\frac{\partial V(D,G)}{\partial\theta_d}$$

② 训练 G fixed D

$$\min_{G}\boxed{V_G}$$

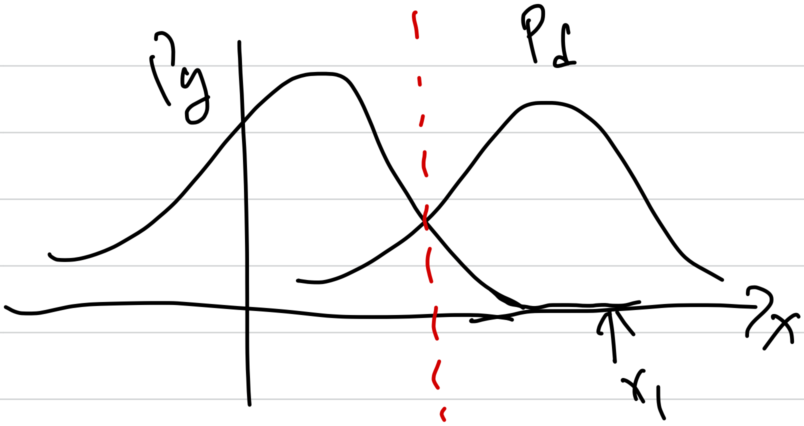
$$\theta_g^{t+1}=\theta^{t}-\eta_g\frac{\partial V_G}{\partial\theta_g}$$

① - ~ -

② - - .

P_g : generated data distribution

P_d : Real 〃 〃 〃



$$V(G,D) = \frac{1}{m}\sum_{i=1}^{m} \log D(x_i) + \frac{1}{m}\sum_{i=1}^{m} \log(1 - D(G(z_i)))$$

$$E_{x \sim p_d} \log D(x) + E_{x \sim p_g} \log(1 - D(x))$$

$m \to \infty$

$$V(G,D) = \int P_d(x) \log D(x)\,dx + \int P_g(x) \log(1 - D(x))\,dx$$

$D^*(x)$ $\forall$ fixed x

$$\tilde{V} = P_d(x) \log D(x) + P_g(x) \log(1 - D(x))$$

$$\frac{\partial \tilde{V}}{\partial D} = 0 \qquad \frac{P_d}{D} + \frac{-P_g}{1 - D} = 0$$

$$\Rightarrow \quad P_d(1 - D) = P_g D$$

$$D = \frac{P_d}{P_g + P_d}$$

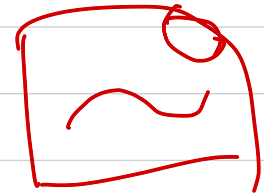
对 $x \sim P_d$ $\quad \boxed{P_d > P_g} \quad D > \frac{1}{2}$

对 $x \sim P_g$ $\quad \boxed{P_d < P_g} \quad D < \frac{1}{2}$

$\Delta u = f$

Proposition 1. $V(G, D)$. max w.r.t. D.

$$D^*(x) = \frac{P_d(x)}{P_d(x)+P_g(x)}$$



Proposition 2. ① 假设 D^* 可以达到.
② 每一步训练 P_g.
都使 V_G 下降 : 存唯一解. $P_g = P_d$

$$V(G, D^*) = \int P_d(x) \log D^*(x)\, dx + \int P_g(x) \log(1 - D^*(x))\, dx$$

$$= \int P_d(x) \log \frac{P_d \cdot \frac{1}{2}}{\frac{P_d+P_g}{2}}\, dx + \int P_g \log \frac{P_g \cdot \frac{1}{2}}{\frac{P_d+P_g}{2}}\, dx$$

$KL\left(P_d \,\|\, \frac{P_d+P_g}{2}\right)$

$KL\left(P_g \,\|\, \frac{P_d+P_g}{2}\right)$

$$= \underline{\int P_d \log \frac{P_d}{\frac{P_d+P_g}{2}} + \int P_g \log \frac{P_g}{\frac{P_d+P_g}{2}}\, dx} + 2\log\frac{1}{2}$$

$$= \underline{J(P_d, P_g)} + 2\log\frac{1}{2}$$

Jensen-Shannon divergence.

$J(P_g, P_d)$: ① cvx in P_g
② $J(P_g, P_d) \geq 0$ "=" iff $P_g = P_d$

§3. Graph Convolutional Network

$X \in R^{n\times d}$ n: 样本数 d: 特征数.

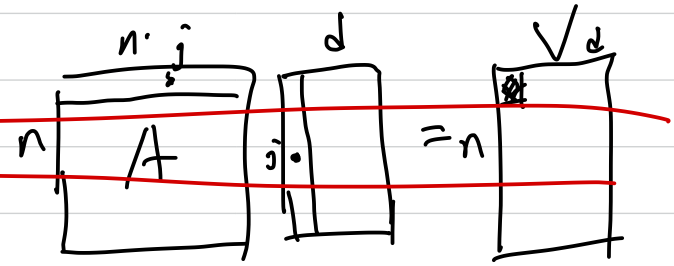
A: adjacent matrix 邻接矩阵.

$A \in R^{n\times n}$ $\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ A_{ij}: jth node to ith node 是否连接.

$i \in S$ Y_i 已知 $i \notin S$ Y_i 未知 要预测

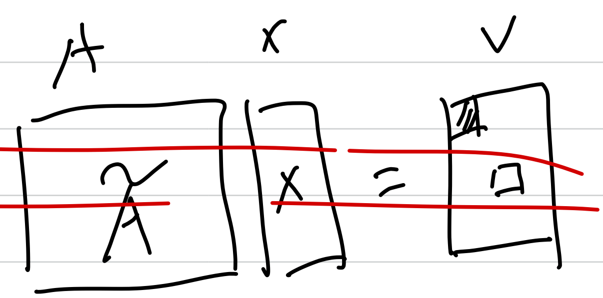
~~$\sigma(XW+b)$~~ $W: R^{d\times m_1}$

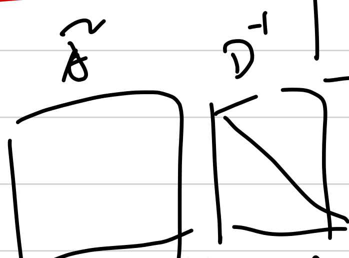
~~AX~~



~~$\tilde{A} = A + I_N$~~

~~$\tilde{A}X$~~





~~V_{ij}: 所有关注 ith 节点~~ 的节点的特征总和

$\tilde{V}_{ij} = A_{ij}/D_{ii}$

$\tilde{A} = A + I_N$

$\tilde{A}\tilde{D}^{-1}X$ $\tilde{D}_{ii} = \sum_{j=1}^{n} A_{ij}$

文章上做：

$$\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}X = V$$

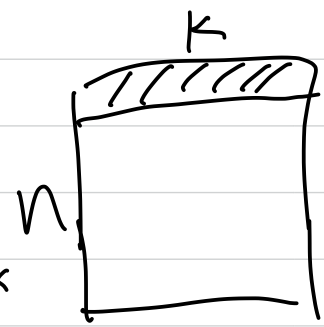
$$H^{(1)} = \sigma(VW^{(0)} + b^{(0)})$$

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)} + b^{(l)})$$

$$H^{(0)} \triangleq X$$

$$H^{L+1} = \mathrm{softmax}(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(L)}W^{(L)} + b^{(L)})$$

$$\boxed{H^{L+1} \in R^{n \times k}}$$



$$\frac{\exp(H^{L+1}_{ij})}{\sum_j \exp(H^{L+1}_{ij})}$$

softmax

$$Z \triangleq H^{L+1}$$

$$L_N = -\sum_{l \in S}\sum_{j=1}^{K} Y_{lj} \ln Z_{lj}$$

$$\theta^{t+1} = \theta^{t} - \eta\frac{\partial L_N}{\partial \theta}$$

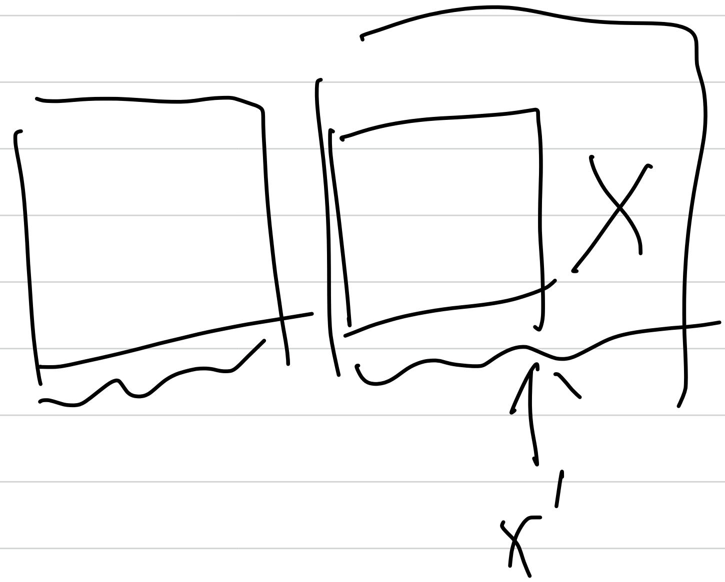
<1> 多层信息

$AX = V$

V行

A^2X

$\boxed{A + A^2 + A^3}$



<2> 为什么是卷积

$\tilde{A}X$