K-Means

Randomly initialize $\rightarrow$ (re)assign $\rightarrow$ mean value

iterate until it converge

Spectral Clustering

to find densely linked Clusters

2 groups A, $B=V\setminus A$, then $cut(A)=\sum_{i\in A, j\notin A} w_{ij}$

Optimize: $$\text{Min} \frac{|\{(i,j)\in E;\ i\in A, j\notin A\}|}{\min\{vol(A),\ 2m-vol(A)\}}$$ ☆ (undirected)

where vol(A): total weight of edges with at least 1 endpoint(s) in A

A: adjacency mat ; $A\vec{x}=\vec{y}$, $y_i=\sum_{(i,j)\in E} x_j$

property: symmetric, Eig vector — real & orthogonal

D: degree mat, 对角阵

L: Laplacian mat $L=D-A$ — The first eig $\lambda=0$

property: Symmetrix, eig $\geq 0 \wedge \in \mathbb{R}$, Eig vec 正交 $\wedge \mathbb{R}$

SVM

to separate data using a line

Maximize the Margin, line: $\vec{w}\cdot\vec{x}+\vec{b}=0$

point $A(x_1,x_2)$, $w=(w_1,w_2)/||w||$

$\gamma_i=(wx_i+b)y_i$

Then goal: $\max \gamma$, $\forall i$, $y_i(wx_i+b)\geq\gamma$

re-define $\gamma=(\frac{w}{||w||}x+\frac{b}{||w||})y$, $|wx+b|=1$ for support vector

Then $\gamma=\frac{1}{||w||}$

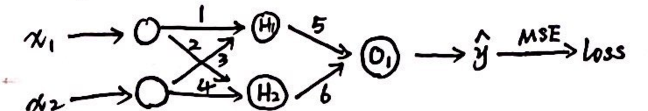
The final goal for hard margin:

$\min_w \frac{1}{2}||w||^2$, $(\forall i, y_i(wx_i+b)\geq 1)$

Fully connect

$\overrightarrow{Output'}=\vec{W}\,\overrightarrow{Input}+\vec{b}$, $\overrightarrow{Output}=f(\overrightarrow{Output'})$

sigmoid $f(z)=\frac{1}{1+e^{-z}}$, $f'(z)=f(z)(1-f(z))$



$E=\frac{1}{2}(y-O_1)^2$, $O_1=f(o_i)$, $o_i=w_5H_1+w_6H_2+b_h$

故 $\frac{\partial E}{\partial w_5}=\underbrace{(O_1-y)f'(o_i)}_{\frac{\partial E}{\partial o_i}\triangleq\delta o_i}\cdot H$, update w_i

$H_1=f(h_1)$, $h_1=w_1x_1+w_3x_2+b$

故 $\frac{\partial E}{\partial w_1}=\frac{\partial E}{\partial H}\cdot\frac{\partial H}{\partial h_1}\cdot\frac{\partial h_1}{\partial w_1}=\delta o_i\cdot w_5\cdot f'(h_1)\cdot x_1$

return $\frac{\partial E}{\partial H}$

Conv NN

$32\times32\times3 \xrightarrow[6\ 5\times5\times3\ \text{filters}]{\text{CONV, RELU}} 28\times28\times6 \longrightarrow \cdots$

$$\text{output size}=\frac{\text{Input}+2\cdot\text{pad}-\text{kernel\_size}}{\text{stride}}+1$$

Numbers of parameters: e.g., 10 5×5 filters for $32\times32\times3$

$N\_p=10\times(5\times5\times3+\underline{1})$ → bias

To conclude,

$W_1\times H_1\times D_1 \xrightarrow[\text{stride } S,\ \text{zero Padding } P]{\text{Filters\_Number } K,\ \text{Spatial extent } F} W_2\times H_2\times D_2$

Where $W_2=\frac{W_1-F+2P}{S}+1$, H_2同理, $D_2=K$

Common settings: K: powers of 2

$(F,S,P)=(3,1,1)/(5,1,2)/(5,2,?)/(1,1,0)$

Architecture

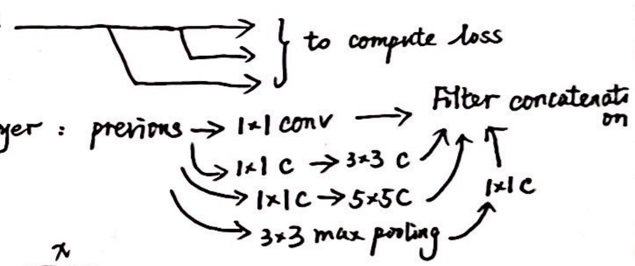
LeNet: CONV → POOL → CONV → POOL → FC → FC → FC
(6 5×5; 2×2 stride 2; 16 5×5; 120; 84; 10)

AlexNet: 双层, 最后交互/在特定位置交互

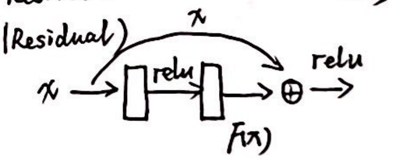
(CONV → MAX POOL → NORM) ×2 → CONV×3 → FC×3

VGGNet: 两个3×3 相当于1个 5×5, 以此类推
更少参数

Inception (Google):



ResNet (Residual):



$H(x)=F(x)+x$

DenseNet — connected to every other layer

KNN

$N_k(x)\triangleq$ 最近的k个 point

$\hat{Y}(x)=\frac{1}{k}\sum_{x_i\in N_k(x)} y_i$, bias ↑ variance ↓

decide k: split data into Train, validation & Test or "fold"s

补充:

solve $(w_0,b_0)=\arg\min ||w||^2$ s.t. $\forall i$, $y_i(\langle w,x_i\rangle+b)\geq 1$

output: $\hat{w}=\frac{w_0}{||w_0||}$, $\hat{b}=\frac{b_0}{||w_0||}$ is a solution of $\max_{||w||=1}\min_i y_i(wx+b)$

proof: (w^*,b^*), $\gamma^*=\min_i y_i(\langle w^*,x_i\rangle+b^*)$

$\Rightarrow y_i(\langle\frac{w^*}{\gamma^*},x_i\rangle+\frac{b^*}{\gamma^*})\geq 1$

Therefore, $||w_0||\leq\frac{1}{\gamma^*}$ which means $\gamma^*\leq\frac{1}{||w_0||}$

$\Rightarrow y_i(\langle\hat{w},x_i\rangle+\hat{b})=\frac{1}{||w_0||}y_i(\langle w_0,x_i\rangle+b_0)\geq\frac{1}{||w_0||}\geq\gamma^*$

$\sum x_i=0$, $\sum x_i^2=1$, $\lambda_2=\frac{x^TLx}{x^Tx}=\sum_{(i,j)\in E}(x_i-x_j)^2$, where x is eig vec

L第2小特征根 → eig vec 可分类.