MAP Estimation

Thursday, June 8, 2017 10:03 AM

DAG - directed acyclic graph

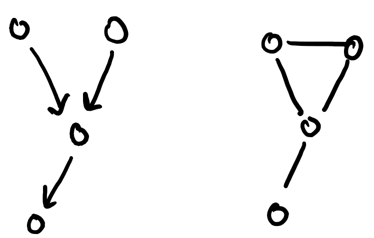
DAG conditional indep is $X_j \perp\!\!\!\perp X_k \mid X_{p_j}$ if (p_j is parent of j)

k is not a descendent of j.

<u>def</u> The <u>moral graph</u> of a DAG has an undirected edge btw X_i, X_j if

(i) $\exists$ a directed edge btw X_j, X_i or

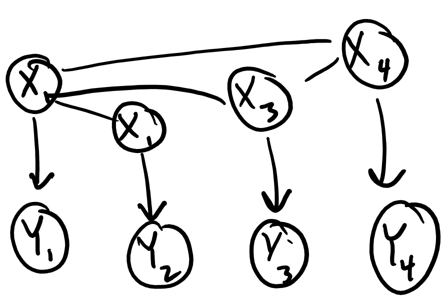
(ii) X_j, X_i are parents of the same node



DAG

$\Rightarrow$ Moral graph has global markov property

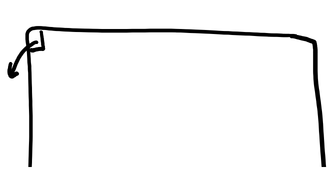
Hidden Markov Models



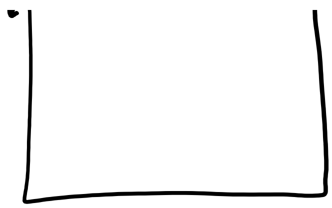
Latent RVs

Observables

use cond. density $f_{X|Y}(x|y)$ to estimate X



x is w/in discrete space



" is w/in discrete space

Maximum A-posteriori estimator (MAP)

$$\hat{x} = \arg\max_{x} f_{X|Y}(x|y) = \arg\max_{x} f_X(x) \cdot f_{Y|X}(y|x)$$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$ — constant in x

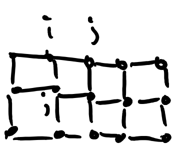
$$-\log f_{Y|X}(y|x) = \frac{1}{n}\sum_{i=1}^{n} \ell(y_i|x_i)$$

$$-\log f_X(x) = \rho(x)$$

MAP: $$\min_{x} \frac{1}{n}\sum_{i=1}^{n} \ell(y_i|x_i) + \rho(x)$$

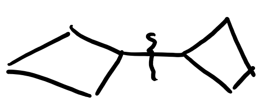
Ising model for $X \in \{-1,1\}^n$

$$f_X(x) = \frac{1}{Z} \exp\left\{ \sum_{i\sim j} W_{ij} x_i x_j + \sum_i a_i x_i \right\}$$

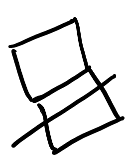


images $\rho(x) \propto \sum_{i\sim j} W_{ij} x_i x_j + \sum_i a_i x_i$

▷ for the Binary case this is solvable as a graph cut problem.



cut = 1

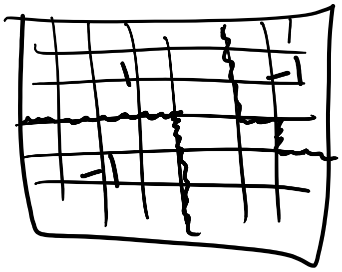


cut = 2

MAP for Ising

$$\min_{x \in \{-1,1\}^n} \frac{1}{n}\sum_i \ell(y_i|x_i) + \sum_{i\sim j} W_{ij} x_i x_j + \sum_i a_i x_i$$

$$= \sum_i b_i x_i + \sum_{i \sim j} W_{ij} x_i x_j$$



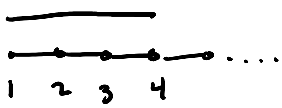
Total Variation denoising

$$\min_{x \in \mathbb{R}^n} \frac{1}{n} \sum_{i=1}^{n} \ell(y_i \mid x_i) + \lambda \sum_{i \sim j} W_{ij} |x_i - x_j|$$

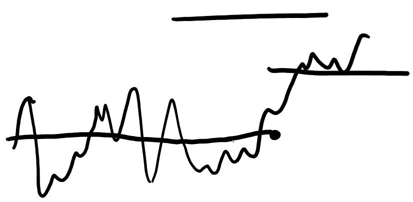
aka fused lasso

for square error loss: $\min_{x} \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i)^2 + \lambda \sum_{i \sim j} W_{ij} |x_i - x_j|$

Can be used for time series ...



(chain graph)



$$W_{ij} = \begin{cases} 1, & |i-j| = 1 \\ 0, & \text{otherwise} \end{cases}$$